

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:16:14 ; Search time 56.55 Seconds  
(without alignments)  
799.417 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEPEEPGRGRGSEFVEMVD.....GPFVTLDMEDCGYNIPQTDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2156	100.0	415	21 AAB07899	Amino acid sequenc
2	2156	100.0	453	21 AAY88438	Modified human asp
3	2156	100.0	453	22 AAEL10642	Human-Asp 2(a) pro
4	2156	100.0	453	22 AAE06872	Human-Asp2(a) delt
5	2156	100.0	453	22 AAU06616	Human-pro-Asp 2(a)
6	2156	100.0	453	22 AAU07215	Human aspartyl pro
7	2156	100.0	453	22 AAE02594	Human-Asp-2(a) del
8	2156	100.0	456	21 AAB07897	Active enzyme port
9	2156	100.0	459	21 AAY88439	Modified human asp
10	2156	100.0	459	22 AAEL10643	Human-Asp 2(a) pro
11	2156	100.0	459	22 AAE06873	Human-Asp2(a) delt

12	2156	100.0	459	22 AAU06617	Human-pro-Asp 2(a)
13	2156	100.0	459	22 AAU07216	Human aspartyl pro
14	2156	100.0	459	22 AAE02595	Human-Asp-2(a) del
15	2156	100.0	460	21 AAB07898	Amino acid sequenc
16	2156	100.0	488	22 AAG66572	Human memapsin 2.
17	2156	100.0	488	22 AAB61334	Memapsin 2 protein
18	2156	100.0	501	21 AAY94767	Human beta-secreta
19	2156	100.0	501	21 AAB07896	Amino acid sequenc
20	2156	100.0	503	22 AAB66573	Human pro-memapsin
21	2156	100.0	503	22 AAB61335	T7 promoter and ve
22	2151	99.8	433	21 AAY88433	Human-pro-Asp-2(a)
23	2151	99.8	433	22 AAEL10640	Human-pro-Asp 2(a)
24	2151	99.8	433	22 AAE06870	Human-pro-Asp2(a)
25	2151	99.8	433	22 AAU06614	Human-pro-Asp 2(a)
26	2151	99.8	433	22 AAU07213	T7-human aspartyl
27	2151	99.8	433	22 AAE02592	Human-pro-Asp-2(a)
28	2151	99.8	446	21 AAY88431	T7-caspase-human-p
29	2151	99.8	446	22 AAEL10638	T7-caspase-human-p
30	2151	99.8	446	22 AAE06868	T7-human-pro-Asp 2
31	2151	99.8	446	22 AAU06612	Human T7-human-pro
32	2151	99.8	446	22 AAU07211	T7-human aspartyl
33	2151	99.8	446	22 AAE02590	T7-human-pro-Asp-2
34	2151	99.8	459	21 AAY88432	T7-caspase-human-p
35	2151	99.8	459	22 AAEL10639	T7-caspase-human-p
36	2151	99.8	459	22 AAE06869	T7-Caspase-Human-p
37	2151	99.8	459	22 AAU06613	Human T7-Caspase-H
38	2151	99.8	459	22 AAU07212	T7-caspase-human a
39	2151	99.8	459	22 AAE02591	T7-caspase-human-p
40	2151	99.8	501	21 AAY88425	Human aspartyl pro
41	2151	99.8	501	22 AAEL10629	Human aspartyl pro
42	2151	99.8	501	22 AAE06859	Human aspartyl pro
43	2151	99.8	501	22 AAU06803	Human aspartyl pro
44	2151	99.8	501	22 AAU07202	Human aspartyl pro
45	2151	99.8	501	22 AAE02581	Human aspartyl pro

ALIGNMENTS

RESULT	1
AAB07899	
ID	AAB07899 standard; Protein; 415 AA.
XX	AC
AC	AAB07899;
XX	AC
DT	14-NOV-2000 (first entry)
XX	
DE	Amino acid sequence of a human beta-secretase enzyme fragment.
XX	
KW	Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
KW	amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
KW	inhibitor.
XX	
OS	Homo sapiens.
XX	
PN	WO200047618-A2.
XX	
PD	17-AUG-2000.
XX	
PF	10-FEB-2000; 2000WO-US03819.
XX	
PR	10-FEB-1999; 99US-0119571.
PR	15-JUN-1999; 99US-0119172.
XX	
PA	(ELAN-) ELAN PHARM INC.
XX	
PI	Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;
PI	Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
XX	WPI; 2000-533011/48.
DR	
XX	Purified beta-secretase protein used in assays to discover inhibitors
PT	which can be used for the treatment of amyloidogenic diseases e.g.

```
PT Alzheimer's disease -
XX
XX Claim 10; Fig 3B; 121pp; English.
XX
XX The specification describes a beta-secretase enzyme. The enzyme cleaves
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
CC enzyme is therefore implicated in the production of amyloid plaque
CC components which accumulate in the brains of individuals afflicted with
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
CC disease-like pathology to test if they maintain or improve cognitive
CC ability or reduce the plaque burden. The compounds are used for the
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC present sequence represents a human beta-secretase enzyme fragment.
XX
XX Sequence 415 AA;
SQ
Query Match 100.0%; Score 2156; DB 21; Length 415;
Best Local Similarity 100.0%; Pred. No. 2.1e-210;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEEPPGRGSGFVEMVDNLRGKSGQGYVEMTVGSPQPTLNILVDTGSSNFVAGAAP 60
Db 1 etdeepeeppgrgsfvemvndlrgksggyvventvgsppqptlnilvdtgssnfavgaap 60
QY 61 HPFLHRYQRLSTYRDLRGVVPYTOGKWEGLCTDLVSIHPGPNVTVRANIAAITE 120
Db 61 hpflhryyqrlslyrdlrgvvyptogkwegelctdlvsihphgpnvtvranaiaite 120
QY 121 SDKFFINGSNWEGILGLAYAEIARPDLSLEPFFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180
Db 121 sdkffingsnwegilglayaeiarpdlslepffdsllvkqthvplnlsfqlcgaagfplnqs 180
QY 181 EVLASVGSMTIIGDHSILYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNVDK 240
Db 181 evlasvgsgmtiigdhsltytgslwytpirrewyvevliivrveingqdlkmdckeynydk 240
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGLFWLGSQLVCWQAGTTPWNIFPVIS 300
Db 241 sivdsgttnlrpkkvfeaaavksikaasstekfpdglfwlgsqlvcwqagttpwnifpvvis 300
QY 301 LYLMEVNTQSFRTIILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
Db 301 lylmgevntqsfritiilpqqylrpvedvatsddcykfaisqsstgtvmgavimegyfvyv 360
QY 361 FDRARKRIGFAVSACHVHDEPRTAAVEGPFVTLDMDCGYNIPQDTE 407
Db 361 fdrarkrigfavsachvhdeirtaavegpfvtldmedcgynipqtd 407
RESULT 2
RAY88438
ID : AAY88438 standard; Protein; 453 AA.
XX
XX AAY88438;
XX
XX 03-AUG-2000 (first entry)
XX
XX Modified human aspartyl protease 2 (Asp2) amino acid sequence.
DE
XX Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
XX Homo sapiens.
OS
XX WO200017369-A2.
PN
XX 30-MAR-2000.
PD
XX
XX 23-SEP-1999; 99WO-US20881.
PF
XX
XX 24-SEP-1998; 98US-0101594.
PR
```

```
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
PI
XX WPI; 2000-303209/26.
DR
XX N-PSDB; AAA15688.
DR
XX
XX New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Example 10; Page 169-172; 183pp; English.
XX
XX This sequence represents a modified human aspartyl protease 2 (Asp2)
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal
CC transmembrane domain deleted. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC nucleotide sequence, and a cell line comprising the vector. Methods for
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
XX Sequence 453 AA;
SQ
Query Match 100.0%; Score 2156; DB 21; Length 453;
Best Local Similarity 100.0%; Pred. No. 2.4e-210;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEEPPGRGSGFVEMVDNLRGKSGQGYVEMTVGSPQPTLNILVDTGSSNFVAGAAP 60
Db 46 etdeepeeppgrgsfvemvndlrgksggyvventvgsppqptlnilvdtgssnfavgaap 105
QY 61 HPFLHRYQRLSTYRDLRGVVPYTOGKWEGLCTDLVSIHPGPNVTVRANIAAITE 120
Db 106 hpflhryyqrlslyrdlrgvvyptogkwegelctdlvsihphgpnvtvranaiaite 165
QY 121 SDKFFINGSNWEGILGLAYAEIARPDLSLEPFFDLSLVKQTHVPLNLSLQLCGAGFPLNQS 180
Db 166 sdkffingsnwegilglayaeiarpdlslepffdsllvkqthvplnlsfqlcgaagfplnqs 225
QY 181 EVLASVGSMTIIGDHSILYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNVDK 240
Db 226 evlasvgsgmtiigdhsltytgslwytpirrewyvevliivrveingqdlkmdckeynydk 285
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGLFWLGSQLVCWQAGTTPWNIFPVIS 300
Db 286 sivdsgttnlrpkkvfeaaavksikaasstekfpdglfwlgsqlvcwqagttpwnifpvvis 345
QY 301 LYLMEVNTQSFRTIILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV 360
Db 346 lylmgevntqsfritiilpqqylrpvedvatsddcykfaisqsstgtvmgavimegyfvyv 405
QY 361 FDRARKRIGFAVSACHVHDEPRTAAVEGPFVTLDMDCGYNIPQDTE 407
Db 406 fdrarkrigfavsachvhdeirtaavegpfvtldmedcgynipqtd 452
RESULT 3
AAE10642
ID : AAE10642 standard; Protein; 453 AA.
XX
XX AAE10642;
```



CC Human aspartyl proteases can act as beta-secretase proteases useful for  
CC treating Alzheimer's disease. App isoforms are useful for identifying  
CC modulators of amyloid-beta peptide production, for use in designing  
CC therapeutics for the treatment and prevention of Alzheimer's disease,  
CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis  
CC and neuronal loss. APP isoforms are also used in methods for identifying  
CC inhibitors and modulators of human Asp2 activity. The invention relates  
CC to a method for identifying agents that modulate the activity of human  
CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
CC as a means to screen in cellular assays for the inhibitors of beta- and  
CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
CC polymerase chain reactions (PCR). The probes are useful for detecting  
CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
CC blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a)  
CC deltaTM protein which is obtained by the deletion of transmembrane domain  
CC at the C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase  
CC activity.

XX SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPGRGSGFVEMVDNLKSGGYYVEMTVGSPQTLLNLYDTGSSNFAVGAAP 60  
Db 46 etdeepepgrgsfvmvndnlrgksgggyvemtvgspqtnllydtgssnfavgaap 105  
QY 61 HPFLHRYQROLSTYRDLRKGVVPTQGWKEGLGDLVSIPHGPNTVVRANIAAITE 120  
Db 106 hpflhryyqrolstydrlrkgyvpytqgkwegelgtlvsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGILGLAYABIAIPDDSLPFFSLVKQTHVPMFLSLQCGAGFFLNQS 180  
Db 166 sdkffingsnwegilglayaeiaipddslpffslvkqthvpmflslqcgagfflnqs 225  
QY 181 EVLASVGSMLIGGTDHSLYSGSLWYTPIRREWYVEVLIIVRVEINGDQDKMCKEYNVDK 240  
Db 226 evlasvgsmliggtldhslysgslwytpirrewyvevliivrveingdqlmckeynydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGQLVCWQAAGTTPWNIFFVIS 300  
Db 286 sivdsgttnlrlpkkvfeaaavksiaasstekfpdgfglwglqvcwqagttppwnlffvis 345  
QY 301 LYLMGEVTVNSFRITLPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 346 lyimgevtvnsfriltpqqylrpvedvatsddcykfaissstgtvmgavimegfyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtd 452

RESULT 5  
AAU06616  
ID AAU06616 standard; Protein; 453 AA.  
XX  
AC AAU06616;  
XX

DT 24-OCT-2001 (first entry)  
XX  
DE Human-pro-Asp 2(a) delta TM.

KW Human; Aspartyl protease; beta-secretase; neurotropic; ASP2;  
KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
KW amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM; mutant; muten.

XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key  
FT Peptide  
1..21  
Location/Qualifiers

FT  
FT Protein /label= Signal\_peptide  
FT 22..453 /label= Mature\_Human\_pro\_Asp\_2(a)\_delta\_TM  
FT Misc-difference 214 /note= "Encoded by CAC"  
XX  
XX WO200149098-A2.  
XX 12-JUL-2001.  
XX  
XX 09-MAY-2001; 2001WO-IB00798.  
XX 09-MAY-2001; 2001WO-IB00798.  
XX (BIEN/) BIENKOWSKI M J.  
XX (GURN/) GURNEY M E. J.  
XX (HEIN/) HEINRIKSON R L.  
XX (PARO/) PARODI L A.  
XX (YANR/) YAN R.  
XX  
XX Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX WPI; 2001-502549/55.  
XX N-PSDB; AAS11530.  
XX  
XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
XX protease 2, lacking Asp2 transmembrane domain and retaining beta  
XX secretase activity of Asp2 useful for identifying inhibitors of Asp2  
XX activity -  
XX  
XX Claim 149; Page 160; 185pp; English.  
XX  
XX The invention relates to a purified polypeptide comprising a fragment of  
XX mammalian aspartyl protease (Asp)2 protein which lacks the Asp2  
XX transmembrane domain and the Asp2 protein, and where the polypeptide and  
XX the fragment retain the beta-secretase activity of the mammalian Asp2  
XX protein. The invention also details polynucleotides for the Asp  
XX proteins and vectors expressing them, and a polypeptide (isoform of  
XX amyloid protein precursor (APP)) comprising the amino acid sequence of an  
XX APP or its fragment containing an APP cleavage site recognizable by a  
XX mammalian beta-secretase, and further comprising two lysine residues at  
XX the carboxyl terminus of the amino acid sequence of the mammalian APP or  
XX APP fragment. Also included in the invention are methods of identifying  
XX modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
XX useful for treating Alzheimer's disease. APP is useful in methods for  
XX identifying inhibitors or modulators of human Asp2 activity and  
XX amyloid-beta (Abeta) peptide production. APP is also useful in designing  
XX therapeutics for the treatment or prevention of Alzheimer's disease.  
XX APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which  
XX is associated with increased levels of Abeta processing is useful in  
XX assays relating the Alzheimer's research. The expression vector is useful  
XX for recombinantly expressing APP. Nucleic acids that hybridize to  
XX Asp oligonucleotides are useful as probes or primers. The probes are  
XX useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
XX Northern and Southern blots. The present sequence is Human-pro-  
XX Asp 2(a) delta TM protein, which lacks the C-terminal transmembrane  
XX domain.  
XX  
XX SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPGRGSGFVEMVDNLKSGGYYVEMTVGSPQTLLNLYDTGSSNFAVGAAP 60  
Db 46 etdeepepgrgsfvmvndnlrgksgggyvemtvgspqtnllydtgssnfavgaap 105  
QY 61 HPFLHRYQROLSTYRDLRKGVVPTQGWKEGLGDLVSIPHGPNTVVRANIAAITE 120  
Db 106 hpflhryyqrolstydrlrkgyvpytqgkwegelgtlvsiphgpnvtvranaiaite 165



QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLPPFFDSLVKQTHVPNLFSLQLCGAGFPNQS 180  
 Db ||||||| 166 skdfingsnwegilglayaeiarpdslpffdsivkqchvnlfsilqlcgagfpnds 225  
 QY 181 EVLASVGSMLIGIDHSLYTGSLWTPPIREWYEVIIIVRVEINGQDLKMDCKEYNDK 240  
 Db ||||||| 226 evlasvgssmligidhsltytgslywtpirweyyevliivveingqdlkmdckeyndk 285  
 QY 241 SIYDSGTTNLRPKKVFEEAAVKSIIKAASSTKEKPPDGFNLGEQLVCHQAGTTPWNIPPVIS 300  
 Db ||||||| 286 siydsdgttnlrpkkveaavksikaasstekfpdgfwlgeqlvcwqagttppwnifpvis 345  
 QY 301 LYLMEVTNQSFRITILPQQYLRPVEDVATSDQDCYKFAISQSSSTGTVMGAVIMEGFYV 360  
 Db ||||||| 346 lylmgevtngsfritilpqylrpvedvatsqdcykfaissstgtvmgavimegfyv 405  
 QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 407  
 Db ||||||| 406 fdrarkrigfavsachvhdefrtaavegpfvtdlmedcgynipqtde 452

## RESULT 6

AAU07215  
 ID AAU07215 standard; Protein; 453 AA.

XX AC AAU07215;

XX DT 24-OCT-2001 (first entry)

XX DE Human aspartyl protease 2a deltaTM (HuAsp-2adeltaTM).

XX KW Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;  
 KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;  
 KW beta-secretase; Alzheimer's disease; HuAsp-2adeltaTM.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..21

FT Protein /note= "Signal peptide"

FT 22..453

XX PN /note= "Mature human aspartyl protease 2a deltaTM"

XX WO200149097-A2.

XX 12-JUL-2001.

XX PF 09-MAY-2001; 2001WO-IB00797.

XX PR 09-MAY-2001; 2001WO-IB00797.

XX PA (BIEN/) BIENKOWSKI M J.

XX PA (GURN/) GURNEY M E.

XX PA (HEIN/) HEINRIKSON R L.

XX PA (PARO/) PARODI L A.

XX PA (YANR/) YAN R.

XX PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;

XX WPI: 2001-502548/55.

XX DR N-PSDB; AAS11715.

XX XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
 PT activity -

XX PS Claim 149; Fig 11; 185pp; English.

XX XX The invention relates to a novel purified polypeptide comprising a  
 CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
 CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
 CC and the fragment retain the beta-secretase activity of the mammalian Asp2

CC protein. Also included is an isoform of amyloid protein precursor (APP)  
 CC comprising the amino acid sequence of a APP or its fragment containing  
 CC an APP cleavage site recognizable by a mammalian beta-secretase, and  
 CC further comprising two lysine residues at the carboxyl terminus of the  
 CC amino acid sequence of the mammalian APP or APP fragment. The  
 CC polypeptides are used for assaying for modulators of beta-secretase  
 CC activity; identifying agents that inhibit the APP processing activity  
 CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that  
 CC modulate the activity of Asp2; and for reducing cellular production of  
 CC amyloid beta (Abeta) from APP. Agents identified by the above methods  
 CC are useful for treating Alzheimer's disease; and for identifying  
 CC modulators of amyloid-beta (Abeta) peptide production, for use in  
 CC designing therapeutics for the treatment or prevention of Alzheimer's  
 CC disease. Probes and primers derived from Asp nucleic acid sequences  
 CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
 CC Northern and Southern blots. The present sequence represents the amino  
 CC acid sequence of human Asp-2a delta TM construct which lacks the  
 CC transmembrane domain. This construct was used for bacterial expression  
 CC and purification of human Asp2a.

XX SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. NO. 2.4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDDEPEEPGRGSGFVEMVDNLRGKSGQGYVEMTVGSPPTNLILVDVIGSSNFAVGAAP 60  
 Db ||||||| 46 etdeepeepgrgsgfvemvdnlrgksggyvemtvgspptnlilvdtgssnfavgaap 105

QY 61 HPFLHRYYQRLSSTYRDLRKGVVYPYTGCKWEGELGTLVSIPIHGPNTVVRANIAAITE 120  
 Db ||||||| 106 hpflhryyqrlsstyrldlrkgvypytgckwegelgtlvsiphgpnvtraniaaite 165

QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLPPFFDSLVKQTHVPNLFSLQLCGAGFPNQS 180  
 Db ||||||| 166 skdfingsnwegilglayaeiarpdslpffdsivkqthvnpnlfsilqlcgagfpnds 225

QY 181 EVLASVGSMLIGIDHSLYTGSLWTPPIREWYEVIIIVRVEINGQDLKMDCKEYNDK 240  
 Db ||||||| 226 evlasvgssmligidhsltytgslywtpirweyyevliivveingqdlkmdckeyndk 285

QY 241 SIYDSGTTNLRPKKVFEEAAVKSIIKAASSTKEKPPDGFNLGEQLVCHQAGTTPWNIPPVIS 300  
 Db ||||||| 286 siydsdgttnlrpkkveaavksikaasstekfpdgfwlgeqlvcwqagttppwnifpvis 345

QY 301 LYLMEVTNQSFRITILPQQYLRPVEDVATSDQDCYKFAISQSSSTGTVMGAVIMEGFYV 360  
 Db ||||||| 346 lylmgevtngsfritilpqylrpvedvatsqdcykfaissstgtvmgavimegfyv 405

QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDTE 407  
 Db ||||||| 406 fdrarkrigfavsachvhdefrtaavegpfvtdlmedcgynipqtde 452

## RESULT 7

AAE02594

ID AAE02594 standard; Protein; 453 AA.

XX AC AAE02594;

XX DT 10-AUG-2001 (first entry)

XX DE Human-Asp-2(a) delta TM protein.

XX KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;  
 KW Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp2a;  
 KW beta-secretase; Asp-2a delta TM.

XX OS Homo sapiens.

XX OS Synthetic.

Key Location/Qualifiers  
 Misc-difference 214  
 /note= "Encoded by CAC"

W0200123533-A2.

05-APR-2001.

22-SEP-2000; 2000WO-US26080.

23-SEP-1999; 99US-0155493.

23-SEP-1999; 99WO-US20881.

13-OCT-1999; 99US-0416901.

06-DEC-1999; 99US-0169232.

(PHAA ) PHARMACIA & UPJOHN CO.

Gurney M, Bienkowski MJ;

WPI: 2001-290516/30.

N-PSDB; RAD06752.

Enzymes that cleave the alpha-secretase site of the amyloid precursor protein, useful for the treatment of Alzheimer's disease -

Example 10; Fig 11; 189pp; English.

The present invention relates to enzymes for cleaving the alpha-secretase site of the amyloid precursor protein (APP) and methods of identifying those enzymes. The methods may be used to identify enzymes that may be used to cleave the alpha-secretase cleavage site of the APP protein. The enzymes may be used to treat or modulate the progress of Alzheimer's disease. The present sequence is human Aspartyl protease 2a (Asp-2a) deltatM protein which is obtained by deleting its transmembrane domain. This sequence has beta-secretase protease activity.

Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAP 60

DB 46 etdepeepgrgrgsfvmvndnlrgskgggyvemtvgspqtlnilvdtgssnfavgaap 105

QY 61 HPFLHRYQRLSSTYRDLRKGYVVPYTGKWEGLGTLVSIPIHGPNTVVRANAAITE 120

DB 106 hpflhryyqrqlsstyrdlrgkgyvpytgkwegeigtldvsiiphgpnvtvranaaite 165

QY 121 SDKFFINGSNWEGILGLAYAEIARPDLSLEPFDSLVKOTHPNLFSLQLCGAGFPLNQS 180

DB 166 skdffingsnwegilglayaeiarpdslslepfdsivkqthvnpnlfsqlcgagfplngs 225

QY 181 EVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDK 240

DB 226 evlasvggsmiiggidhsltygslwtptirrewyeviiivrveingqdlkmdckeynydk 285

QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVIS 300

DB 286 sivdsgttnlrpkkvfeaaavksiaasstekfpdgfglgeqlvcwqagtppwnipfvvis 345

QY 301 LYLMEVTNQSFRITILPQOYLREVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVV 360

DB 346 lylmevtngsfrtilpqqylrpvedvatsqddcykfaissgstgtvmgavimegyfyyv 405

QY 361 FDRARKRIGFVAVSCHVHDEFRTAAVGGPFVTLDMEDCGYNIPQDTE 407

DB 406 fdrarkrigrfavsachvhdefirtaavgeppfvtlidmedcgynipqtde 452

RESULT 8

AAB07897

ID AAB07897 standard; Protein; 456 AA.

XX AC AAB07897;

XX DT 14-NOV-2000 (first entry)

XX DE Active enzyme portion of human beta-secretase enzyme.

XX KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide; amyloid plaque component; Alzheimer's disease; amyloidogenic disease; inhibitor; ss.

XX OS Homo sapiens.

XX PN W0200047618-A2.

XX PD 17-AUG-2000.

XX PF 10-FEB-2000; 2000WO-US03819.

XX PR 10-FEB-1999; 99US-0119571.

XX PR 15-JUN-1999; 99US-0139172.

XX PA (ELAN-) ELAN PHARM INC.

XX PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

XX PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX DR WPI; 2000-533011/48.

XX PT Purified beta-secretase protein used in assays to discover inhibitors which can be used for the treatment of amyloidogenic diseases e.g. Alzheimer's disease -

XX PS Claim 24; Fig 2B; 121pp; English.

XX CC The specification describes a beta-secretase enzyme. The enzyme cleaves beta-amyloid precursor protein to produce beta-amyloid peptide. This enzyme is therefore implicated in the production of amyloid plaque components which accumulate in the brains of individuals afflicted with Alzheimer's disease. Inhibitors of beta-secretase are administered to a mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-like pathology to test if they maintain or improve cognitive ability or reduce the plaque burden. The compounds are used for the treatment of amyloidogenic diseases e.g. Alzheimer's disease. The present sequence represents the active enzyme portion of human beta-secretase enzyme.

XX SQ Sequence 456 AA;

Query Match 100.0%; Score 2156; DB 21; Length 456;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAP 60

DB 1 etdepeepgrgrgsfvmvndnlrgskgggyvemtvgspqtlnilvdtgssnfavgaap 60

QY 61 HPFLHRYQRLSSTYRDLRKGYVVPYTGKWEGLGTLVSIPIHGPNTVVRANAAITE 120

DB 61 hpflhryyqrqlsstyrdlrgkgyvpytgkwegeigtldvsiiphgpnvtvranaaite 120

QY 121 SDKFFINGSNWEGILGLAYAEIARPDLSLEPFDSLVKOTHPNLFSLQLCGAGFPLNQS 180

DB 121 skdffingsnwegilglayaeiarpdslslepfdsivkqthvnpnlfsqlcgagfplngs 180

QY 181 EVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDK 240

DB 181 evlasvggsmiiggidhsltygslwtptirrewyeviiivrveingqdlkmdckeynydk 240

QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVIS 300

Db 241 sldvsgttnlrpkvfeavkskaasstekfpgfwlgeqlvcwgaattppwnifpvis 300  
QY 301 LVLGMGVNQSPRITILPQOYLPRVEDVATSDDCYKFAISOSSCTGVMGAVIMEGFYV 360  
Db 301 lylmgevtnqsfritilpqgylrpvedvatsqddcykfaissstgtvmgavimegfyyv 360  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 407  
Db 361 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgylnipqtde 407  
RESULT 9  
AAE10643 9  
ID AAY08439 standard; Protein; 459 AA.  
AC AAY88439;  
XX  
DT 03-AUG-2000 (first entry)  
XX  
DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.  
XX  
KW Aspartyl protease; aspartase; amyloid precursor protein; App; Asp 2;  
KW Alzheimer's disease; beta secretase site.  
XX  
OS Homo sapiens.  
XX  
PN WO200017369-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 23-SEP-1999; 99WO-US20881.  
XX  
PR 24-SEP-1998; 98US-0101594.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;  
XX  
DR WPI; 2000-303209/26.  
DR N-PSDB; AAA15689.  
XX  
PT New enzyme designated human aspartase useful in research into  
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at  
PT the beta secretase site to produce amyloid beta peptide -  
XX  
PS Example 10; Page 173-176; 183pp; English.  
XX  
CC This sequence represents a modified human aspartyl protease 2 (Asp2)  
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal  
CC transmembrane domain deleted. The invention relates to a protease  
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid  
CC precursor protein (APP). The protease contains a sequence encoding the  
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by  
CC 100-300 amino acids. When mutated the APP gene causes an autosomal  
CC dominant form of Alzheimer's disease. APP localises to the cell surface  
CC membrane and have a single C-terminal transmembrane domain. Proteolytic  
CC processing of APP produces the amyloid beta protein, which is possibly  
CC very important in Alzheimer's disease. The invention includes a  
CC nucleotide sequence encoding the protease, a vector containing the  
CC nucleotide sequence, and a cell line comprising the vector. Methods for  
CC screening for inhibitors of beta secretase activity are also given in the  
CC invention. The human aspartase protein and nucleotide sequences and the  
CC methods for identifying inhibitors of the protease, are useful in the  
CC treatment of and research in to Alzheimer's disease.  
XX  
SQ Sequence 459 AA;

Query Match 100.0%; Score 2156; DB 21; Length 459;  
Best Local Similarity 100.0%; Pred. NO. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPEGRGSGSTEMVDNLRKSGCGYXVEMTVGSPQTNLILVDTGSSNFAVGAAP 60  
Db etdepepegrgsgstemvdlrkgsggyvemtvgspqtnlilvdtgssnfavgaap 105  
QY 61 HPFLHRYVQROLSSPYRDLRKGVVYPYTGQKWEGLGTDLVSIHPGPNVTVRANIAAITE 120  
Db hpflhryvqrolsspyrdlrkgvvyptqgkweglgtdlvsihpgpnvtvranaiaite 165  
QY 121 SDKFFINGSNMEGILGLAYAEIARPDSDLEPFFDSLVKQTHVPLNLSLQLCGAGFPLNQS 180  
Db sdkffingsnwegilglayaeiarpddslepfdslnkqthvplnlsfqlcgaagfplnqs 225  
QY 181 EVLASVCGSMIIGIDHSLYTGSLWYPIRREWYVEVTVIRVEINGODLKMDCKEYNYDK 240  
Db evlasvsgsmilgldhslytgslwytpirrewyvevlivrveingdalkmdckeynydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLVQWAGTTPWNIFPVIS 300  
Db sivdsgttnlrpkkvfeavksikaasstekfpgfwlgeqlvcwgaattppwnifpvis 345  
QY 301 LYLGMGVNQSPRITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db lylmgevtnqsfritilpqgylrpvedvatsqddcykfaissstgtvmgavimegfyyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 407  
Db fdrarkrigfavsachvhdefrtaavegpfvtldmedcgylnipqtde 452

RESULT 10  
AAE10643  
ID AAE10643 standard; Protein; 459 AA.  
XX  
AC AAE10643;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE Human-Asp 2(a) protein with (His)6 tag and lacking TM domain.  
XX  
KW Human; aspartyl protease 2a; Asp2a; amyloid precursor protein; APP;  
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;  
KW amyloid plaque; neuronal loss; proteolytic; neurotrophic; neuroprotective.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..459  
FT /note= "Mature Human-Asp2(a) deltaTM (His)6 protein"  
FT Misc-difference 214 /note= "Encoded by CAC"  
FT Misc-difference 454 /note= "Encoded by CAG"  
FT Misc-difference 455 /note= "Encoded by CAG"  
FT Misc-difference 456 /note= "Encoded by CAG"  
FT Misc-difference 457 /note= "Encoded by CAG"  
FT Misc-difference 458 /note= "Encoded by CAG"  
FT Misc-difference 459 /note= "Encoded by CAG"  
FT /note= "Encoded by CAG"  
PN GB2357767-A.  
XX  
XX 04-JUL-2001.  
PD  
XX 22-SEP-2000; 2000GB-0023315.  
PF  
XX 23-SEP-1999; 99US-0155493.  
PR

[illegible]

modulators of amyloid-beta peptide production, for use in designing therapeutics for the treatment and prevention of Alzheimer's disease, dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis and neuronal loss. APP isoforms are also used in methods for identifying inhibitors and modulators of human Asp2 activity. The invention relates to a method for identifying agents that modulate the activity of human aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used as a means to screen in cellular assays for the inhibitors of beta- and gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in polymerase chain reactions (PCR). The probes are useful for detecting Hu-Asp nucleic acids in in vitro assays and in Northern and Southern blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a) deltaTM (His)6 protein which is obtained by deletion of C-terminal transmembrane domain and addition of a hexa-Histidine tag at the C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase activity.

XX Sequence 459 AA;

Query Match 100.0%; Score 2156; DB 22; Length 459;  
Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPEPEGRGRGSFVEMVNLKCKSGGYVEMTVGSPPTLNLVDTGSSNFAVGAAP 60  
AAU06617  
Db 46 etdeepepgrgrgsfvmvndlrgsksggyvemtvgspptlnllydtgssnfavgaa 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVVPTQGWKEGELGTLVSIPIHGNVTVRANIAITE 120  
Db 106 hpflhryqrlsystyrdlrgvvyptqgkwegeltvlsviphgnvtvranaiaite 165  
QY 121 SDKEFFINGSNWEGILGAYAEIARPDLSLEFFDLSLVKQTHVPNLFSLQLCGAGFPLNQS 180  
Db 166 sdkffingsnwegilgayaearpdsleffdsllvkqthvnlfsllqlcgaafplnqs 225  
QY 181 EVLASVGSMTIGIDHSHTGSLWYTPIRRETYEVIIVRVEINGDLKMDCKEYNKYD 240  
Db 226 evlasvgsmtlgiidhshtgslwytpirrewyeviivrveingdldkmdckeynydk 285  
QY 241 SIYDGSNTNLRPKKVFEEAVKSIKAASSTKFPDGFGLGELVCGWAGTTPWNIFPVIS 300  
Db 286 siydsntnlrlpkkvfeavksikaasstekfpdgfwlgeqlvcwaggttwnifpv 345  
QY 301 LYLMEVTNSFRITILPOQLRPVEDVATSQDCYKFAISQSTGTVMGAVIMEGYVV 360  
Db 346 lylmgevtndsfritilpqqlrpvedvatsqdcykfaissqstgtvmgavimegyvv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVGPFTLDMEDCGYNIPQTD 407  
Db 406 fdrarkrigfavsachvhdefrtaaavgpftldmedcgynipqtde 452

RESULT 12  
AAU06617

ID AAU06617 standard; Protein; 459 AA.

AC AAU06617;

XX 24-OCT-2001 (first entry)

DT Human-pro-Asp 2(a) delta TM (His)6.

DE Human; Aspartyl protease; beta-secretase; neurotropic; ASP2;  
KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
KW amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM (His)6; mutant; muten.  
XX Homo sapiens.  
OS Synthetic.

XX Key Location/Qualifiers  
FH 1..21  
FT Peptide /label= Signal\_peptide  
FT Protein 22..459

FT Misc-difference 214 /label= Mature\_Human\_pro\_Asp\_2(a)\_delta\_TM\_(His)6  
FT /note= "Encoded by CAC"  
FT Misc-difference 454..459  
FT /note= "Encoded by CAGCAGCAGCAGCAGCAG"  
FT Region 454..459  
FT /label= His\_tag  
FT /note= "Nickel binding region to aid purification"

XX WO200149098-A2.

XX 12-JUL-2001.

XX 09-MAY-2001; 2001WO-IB00798.

XX 09-MAY-2001; 2001WO-IB00798.

XX (BIEN/) BIENKOWSKI M J.

XX (GURN/) GURNEY M E.

XX (HEIN/) HEINRIKSON R L.

XX (PARO/) PARODI L A.

XX (YANR/) YAN R.

XX Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;

XX WPI; 2001-502549/55.

XX N-PSDB; AAS11531.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
FT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity

XX Claim 149; Fig 12; 185pp; English.

XX The invention relates to a purified polypeptide comprising a fragment of  
CC mammalian aspartyl protease (Asp)2 protein which lacks the Asp2  
CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
CC the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. The invention also details polynucleotides for the Asp  
CC proteins and vectors expressing them, and a polypeptide (isoform of  
CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
CC APP or its fragment containing an APP cleavage site recognizable by a  
CC mammalian beta-secretase, and further comprising two lysine residues at  
CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
CC APP fragment. Also included in the invention are methods of identifying  
CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
CC useful for treating Alzheimer's disease. APP is useful in methods for  
CC identifying inhibitors or modulators of human Asp2 activity and  
CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
CC therapeutics for the treatment or prevention of Alzheimer's disease.  
CC APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which  
CC is associated with increased levels of Abeta processing is useful in  
CC assays relating the Alzheimer's research. The expression vector is useful  
CC for recombinantly expressing APP. Nucleic acids that hybridize to  
CC Asp oligonucleotides are useful as probes or primers. The probes are  
CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence is Human-pro-  
CC Asp 2(a) delta TM (His)6 protein, which lacks the C-terminal  
CC transmembrane domain and has a His tag to aid purification.

XX Sequence 459 AA;

Query Match 100.0%; Score 2156; DB 22; Length 459;  
Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDPEPEGRGRGSFVEMVNLKCKSGGYVEMTVGSPPTLNLVDTGSSNFAVGAAP 60  
Db 46 etdeepepgrgrgsfvmvndlrgsksggyvemtvgspptlnllydtgssnfavgaa 105  
QY 61 HPFLHRYQRLSSTYRDLRKGVVPTQGWKEGELGTLVSIPIHGNVTVRANIAITE 120

Db	106	hpfihryyqrlsstyrldlrkgyvpytgkgwgelgtdlvsiphgnvtraniaaite	165
QY	121	SDKFFINGSNWEGILGLAYAEIARPDLSLEPFDSLKQTHVPLNLSLQICGAGFPLNQS	180
Db	166	sdkfingsnwegilglayaeiarpdlslepfdslvkqthvplnlsqicgagfplngs	225
QY	181	EVLASVGSMTIIGIDHSLYTGSLWTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDK	240
Db	226	evlasvggsmiigidhslytgslytbpirrewyevliivrveingqdlkmdckeynydk	285
QY	241	SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNIPFVIS	300
Db	286	sivdsgttnlrpkkvfeaaavksikaasstekfpgdglvgeqlvcwaggttwnipfvvis	345
QY	301	LYLMGEVTNQSFRITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV	360
Db	346	lylmgevtnqsfrilitpqgylrpvedvatsqddcykfaissgstgtvmgavimegyfvv	405
QY	361	FDRAKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407	
Db	406	fdarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtde 452	
RESULT 13			
AAU07216			
ID	AAU07216	standard; Protein; 459 AA.	
XX	AAU07216;		
XX	24-OCT-2001	(first entry)	
DT	XX	Human aspartyl protease 2a deltaTM (His)6.	
DE	XX	Human; aspartyl protease 1; Asp-1; neutropic; neuroprotective;	
KW	XX	aspartyl protease 2; Asp2; amyloid protein precursor; APP;	
KW	XX	beta-secretase; Alzheimer's disease; HuAsp-2adeltaTM (His)6.	
OS	XX	Homo sapiens.	
XX	Key	Location/Qualifiers	
FT	Peptide	1..21 "Signal peptide"	
FT	/note=	22..459	
FT	Protein	/note= "Mature human aspartyl protease 2a deltaTM (His)6"	
FT	/note=	454	
FT	Misc-difference	/note= "Encoded by cag"	
FT	/note=	455	
FT	Misc-difference	/note= "Encoded by cag"	
FT	/note=	456	
FT	Misc-difference	/note= "Encoded by cag"	
FT	/note=	457	
FT	Misc-difference	/note= "Encoded by cag"	
FT	/note=	458	
FT	Misc-difference	/note= "Encoded by cag"	
FT	/note=	459	
FT	Misc-difference	/note= "Encoded by cag"	
XX	WO200149097-A2.		
PN	XX	12-JUL-2001.	
XX	09-MAY-2001;	2001WO-IB00797.	
PF	XX	09-MAY-2001;	2001WO-IB00797.
XX	XX	(BIEN/) BIENKOWSKI M J.	
PA	XX	(GURN/) GURNEY M E.	
PA	XX	(HEIN/) HEINRIKSON R L.	
PA	XX	(PARO/) PARODI L A.	
XX	XX	(YANR/) YAN R.	
PI	XX	Bienskowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;	

XX	WPI; 2001-502548/55.		
DR	N-PSDB; AAS11716.		
XX	Novel purified polypeptide comprising fragment of mammalian aspartyl		
PT	protease 2, lacking Asp2 transmembrane domain and retaining beta		
PT	secretase activity of Asp2 useful for identifying inhibitors of Asp2		
PT	activity		
XX	Claim 149; Fig 12; 185pp; English.		
PS	XX	The invention relates to a novel purified polypeptide comprising a	
XX	CC	fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the	
CC	CC	Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide	
CC	CC	and the fragment retain the beta-secretase activity of the mammalian Asp2	
CC	CC	protein. Also included is an isoform of amyloid protein precursor (APP)	
CC	CC	comprising the amino acid sequence of a APP or its fragment containing	
CC	CC	an APP cleavage site recognisable by a mammalian beta-secretase, and	
CC	CC	further comprising two lysine residues at the carboxyl terminus of the	
CC	CC	amino acid sequence of the mammalian APP or APP fragment. The	
CC	CC	polypeptides are used for assaying for modulators of beta-secretase	
CC	CC	activity; identifying agents that inhibit the APP processing activity	
CC	CC	of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that	
CC	CC	modulate the activity of Asp2; and for reducing cellular production of	
CC	CC	amyloid beta (Abeta) from APP. Agents identified by the above methods	
CC	CC	are useful for treating Alzheimer's disease; and for identifying	
CC	CC	modulators of amyloid-beta (Abeta) peptide production, for use in	
CC	CC	designing therapeutics for the treatment or prevention of Alzheimer's	
CC	CC	disease. Probes and primers derived from Asp nucleic acid sequences	
CC	CC	are useful for detecting Hu-Asp nucleic acids in in vitro assays and in	
CC	CC	Northern and Southern blots. The present sequence represents the amino	
CC	CC	acid sequence of human Asp-2a delta TM (His)6 construct which has	
CC	CC	a 6 histidine tag and lacks the transmembrane domain. This construct was	
CC	CC	used for expression and purification of human Asp2a in insect cells.	
XX	Sequence	459 AA;	
SQ	Query Match	100.0%; Score 2156; DB 22; Length 459;	
	Best Local Similarity	100.0%; Pred. No. 2.5e-210;	
	Matches 407; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ETDEEPEEPRGRGSFVEMVDNLKCGSGQGYVVTQGWEGELGTDLSVIPHGNVTVRANIAAITE	60
Db	46	etdeeepeepgrrgsfvfmvndnlrgksqgyvvtvtpqgkwegetldlvsiphgnvtraniaaap	105
QY	61	HPFLHRYYQRLSSTYRDLRKGVVYPTQGWEGELGTDLSVIPHGNVTVRANIAAITE	120
Db	106	hpfihryyqrlsstyrldlrkgyvpytgkgwgelgtdlvsiphgnvtraniaaite	165
QY	121	SDKFFINGSNWEGILGLAYAEIARPDLSLEPFDSLKQTHVPLNLSLQICGAGFPLNQS	180
Db	166	sdkfingsnwegilglayaeiarpdlslepfdslvkqthvplnlsqicgagfplngs	225
QY	181	EVLASVGSMTIIGIDHSLYTGSLWTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDK	240
Db	226	evlasvggsmiigidhslytgslytbpirrewyevliivrveingqdlkmdckeynydk	285
QY	241	SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNIPFVIS	300
Db	286	sivdsgttnlrpkkvfeaaavksikaasstekfpgdglvgeqlvcwaggttwnipfvvis	345
QY	301	LYLMGEVTNQSFRITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVV	360
Db	346	lylmgevtnqsfrilitpqgylrpvedvatsqddcykfaissgstgtvmgavimegyfvv	405
QY	361	FDRAKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407	
Db	406	fdarkrigfavsachvhdefrtaavegpfvtldmedcgynipqtde 452	
RESULT 14			
AAE02595			



**us-09-724-571-58.rag**

CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme fragment.

CC	present sequence
XX	
SO	Sequence 460 AA;

Query Match	100.0%;	Score 2156;	DB 21;	Length 460;
Best Local Similarity	100.0%;	Pred. No. 2.5e-210;		
Matches 407;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ETDEPEEPGRGSFVEMDNLRGKSGQGYVEMTVGSPQOTLNLIVDTGSSNEAVGAAP	60	
Db	46	etdeepeepgrgsfvmvdmnlrgksgqgyyvemtvgspqotlnilvdtgssnfavgaaap	105	
QY	61	HPFLHRYQRQLSSTYRDLRKGVVYPYPTQGWEGELGTDLVSIHPGPNVTVRANIAAITE	120	
Db	106	hpfllhryyqrqlsstyrdlrgkvvypytqgkwegelgtdlvsihpgpnvtraniaaite	165	
QY	121	SDKFFINGSNWEGLILGAYAEIARPDDSLEPFPSLVKQTHVPNLFSQLQCGAGPPLNQS	180	
Db	166	sdkffingsnwegilgayaearpddslepfpslvkqthvpnlfsqlqcgagfplnqs	225	
QY	181	EVLASVGSMIIGGDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK	240	
Db	226	evlasvgsmiiggdhsllytgslywtpirrewyveiivrveingqdlkmdckeynydk	285	
QY	241	SIVDSGTTNLRPLPKVYFAAOKSIKAAASSTKEPDPGFWLGQOLVCQWAGTTPWNFFPVIS	300	
Db	286	sivdsgettlnlrplpkvfyfaaoksiikaasstkepdpdgfwlgqolvcwagttcpwnffpv	345	
QY	301	LYLMGEVNTQSFRTILPQQYLRPVEDVATSDCCYKFAISQSSGTGTVMGAVIMEGFYV	360	
Db	346	lylmgevntqsfritlpqqylrpvedvatqdcykyaisqssgtgtvmgavimegfyvv	405	
QY	361	FDRARKRTGFAVSACHVIDEPTAAVEGPFVTLDMEDCGYNIPQDTE	407	
Db	406	fararkrtlgfavsachvndeptaavegpfvtldmedcgyinlpqdtte	452	

Search completed: August 7, 2002, 09:16:15  
Job time: 160 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:14:17 ; Search time 32.31 Seconds  
(without alignments)  
1210.410 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEPEPGRGSGFVEMVD.....GPFVLDMDGYNIPQTD 407

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2156	100.0	501	2 A59090	aspartic proteinase
2	308	14.3	384	2 JC7574	pepsinogen A - Afr
3	308	14.3	385	2 JC7575	pepsinogen A - bul
4	306	14.2	387	2 B38302	pepsin (EC 3.4.23)
5	303	14.1	383	2 JC7573	pepsinogen C - Afr
6	302.5	14.0	388	1 S19682	pepsin A (EC 3.4.23)
7	301	14.0	382	1 PECH	pepsin A (EC 3.4.23)
8	299.5	13.9	396	2 A34401	cathepsin E (EC 3.4.23)
9	298.5	13.8	383	2 A41443	pepsin (EC 3.4.23)
10	298.5	13.8	384	2 A39314	gastricsin (EC 3.4.23)
11	296	13.7	387	2 B38302	pepsin (EC 3.4.23)
12	296	13.7	391	2 A43356	cathepsin E (EC 3.4.23)
13	295.5	13.7	412	1 KHUD	cathepsin D (EC 3.4.23)
14	295	13.7	387	2 D38302	pepsin (EC 3.4.23)
15	289.5	13.4	388	1 S19684	pepsin A (EC 3.4.23)
16	287.5	13.3	444	2 T24204	hypothetical prote
17	287	13.3	407	1 KHRD	cathepsin D (EC 3.4.23)
18	286	13.3	387	2 E38302	pepsin (EC 3.4.23)
19	285.5	13.2	398	2 S66465	cathepsin E (EC 3.4.23)
20	285	13.2	398	2 I51185	cathepsin D (EC 3.4.23)
21	283.5	13.1	388	1 PEMQAR	pepsin A (EC 3.4.23)
22	282.5	13.1	388	1 PMSH	pepsin A (EC 3.4.23)
23	281.5	13.1	381	1 CSHB	chymosin (EC 3.4.23)
24	281.5	13.1	388	2 PEMQAJ	pepsin A (EC 3.4.23)
25	281.5	13.1	388	2 A30142	pepsin A (EC 3.4.23)
26	279.5	13.0	388	2 B30142	pepsin A (EC 3.4.23)
27	279.5	13.0	410	1 KMSD	cathepsin D (EC 3.4.23)
28	278.5	12.9	385	1 PEFG	pepsin A (EC 3.4.23)
29	278	12.9	387	2 JC7245	pepsinogen A - com

30 277.5 12.9 380 2 I47176 chymosin (EC 3.4.23)  
31 277.5 12.9 396 2 S36865 cathepsin E (EC 3.4.23)  
32 276 12.8 389 2 JE0371 pepsin C (EC 3.4.23)  
33 273.5 12.7 381 1 CMBO chymosin (EC 3.4.23)  
34 270.5 12.5 377 1 PEMQJ gastricsin (EC 3.4.23)  
35 270.5 12.5 389 2 A38302 pepsin (EC 3.4.23)  
36 270 12.5 376 2 I45856 aspartic proteinase  
37 268.5 12.5 344 1 KHPGD cathepsin D (EC 3.4.23)  
38 267.5 12.4 381 2 JC7247 prochymosin - com  
39 266 12.3 380 2 S03433 candidapepsin (EC 3.4.23)  
40 266 12.3 405 2 A25379 saccharopepsin (EC 3.4.23)  
41 264 12.2 396 2 T47207 aspartic proteinase  
42 263.5 12.2 388 2 JC7246 pepsinogen C - com  
43 262.5 12.2 394 2 B43356 gastricsin (EC 3.4.23)  
44 261.5 12.1 387 2 A45117 aspartic proteinase  
45 261.5 12.1 388 2 A29937 gastricsin (EC 3.4.23)

#### ALIGNMENTS

RESULT 1

A59090  
aspartic proteinase (EC 3.4.23.-) BACE precursor - human  
N:Alternate names: beta-secretase; beta-site APP cleaving enzyme  
C:Species: Homo sapiens (man)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-May-2000  
C:Accession: A59090  
M:A.; B.: Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Tople  
Science 286, 735-741, 1999  
A:Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran  
A:Reference number: A59090; MUID:20002972  
A:Note: submitted to GenBank, September 1999  
A:Accession: A59090  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-501 <VAS>  
A:Cross-references: GB:AF190725; NID:96118538; PIDN:AAF04142.1; PID:96118539  
C:Genetics:  
C:Superfamily: beta-secretase  
C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-45/Domain: propeptide #status predicted <PRO>  
F:46-501/Product: acid proteinase BACE #status predicted <MAT>  
F:461-477/Domain: transmembrane #status predicted <TRN>  
F:93,289/Active site: Asp #status predicted  
F:153,172,223,354/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2156; DB 2; Length 501;  
Best Local Similarity 100.0%; Pred. No. 1.6e-173;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ETDEPEPGRGSGFVEMVDNLKSGQGYVEMTVGSPPTNLNLDVTGSSNFAVGAAP 60  
|||||  
Db 46 ETDEPEPGRGSGFVEMVDNLKSGQGYVEMTVGSPPTNLNLDVTGSSNFAVGAAP 105  
|||||  
Qy 61 HPFLHRYQRLSTYRDLRGVYVPTQGWEGELGTDLSVPHGPNVTVRANIAITE 120  
|||||  
Db 106 HPFLHRYQRLSTYRDLRGVYVPTQGWEGELGTDLSVPHGPNVTVRANIAITE 165  
|||||  
Qy 121 SDKFFINGSNWEGTLGLAYAEIAPDLSLPPFDLSLVKQTHVPLNLSLQCGAGFPUNQS 180  
|||||  
Db 166 SDKFFINGSNWEGTLGLAYAEIAPDLSLPPFDLSLVKQTHVPLNLSLQCGAGFPUNQS 225  
|||||  
Qy 181 EVLASVGSMTIGIDHSLTGTSLWYTPIRREWYEVIIIVRVEINGDLMCKCKEYNDK 240  
|||||  
Db 226 EVLASVGSMTIGIDHSLTGTSLWYTPIRREWYEVIIIVRVEINGDLMCKCKEYNDK 285  
|||||  
Qy 241 SIYDSGTNLRPKVFEAAVKSIKAASSTKEKPDGFWLGEQLVCWQAGTTPWNIFPVIS 300  
|||||

J. Biochem. 129, 147-153, 2001  
 A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinog  
 A:Reference number: JC7573; MUID:21064922; PMID:11134969  
 A:Contents: Stomach  
 A:Accession: JC7575  
 A:Molecule type: mRNA  
 A:Residues: 1-385 <IKU>  
 A:Cross-references: DDBJ:AB045376  
 A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like  
 C:Genetics:  
 A:Gene: Pga  
 C:Superfamily: pepsin  
 C:Keywords: stomach; zymogen

Query Match 14.3%; Score 308; DB 2; Length 385;  
 Best Local Similarity 27.8%; Pred. No. 4.8e-18;  
 Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVG---AAPHFLHRYQRLSTYRDLRKGYYV 85  
 73 YFTGISIGTPPQSFVTFIDFGSSNLWVPSVYCSNPACTNHHMFNPQOSTFQATNTPVSI 132  
 86 PYTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAEIAR 144  
 133 QYGTGSMGFLGYDTQV---GNIQTNIQIFGLSQSEPGSFLYSPFDGILGLAFPSLA- 188  
 145 PDSLEPFDLSVKQTHVP--NLFSLQLCGAGFPINQSEVLASVGGSMIIGIDHSLYTGS 203  
 189 -SSQATPVFDNMWQGLIPQDLFSVYLSOG---QS-----GSFVFGVDTSYITGN 237  
 204 LWTPIRREWYEVIIIVRVEINGDQLK--DKKEYNYDKSIVDSGTTNLRPKKVFEEAV 261  
 238 LNWPLTAETHYWOITVDSISIGGOVIACSGSC-----SAIVDTGTSLLAGP----STPI 287  
 262 KSIKAASSTEKFPDGFWMGEOLVQWQAGTTPWNTFFVLSYLMGEVINOSEFRITILPQOY 321  
 288 ANIQYIGANODSNGQYV---INCNNISNMPTVVF-----TINGVOY 326

Query Match 14.3%; Score 308; DB 2; Length 384;  
 Best Local Similarity 27.5%; Pred. No. 4.8e-18;  
 Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL-----HRYQRLSTYRDLRKG 82  
 72 YGTGISIGTPPQSFVTFIDFGSSNLWV---PSVYCSQACSNHNFNPQOSTFQATNTP 128  
 83 YVPTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAE 141  
 129 VSIQYGTGSMGFLGYDTQV---GNIQISNQMFGLSESEPGSFLYSPFDGILGLAFPS 185  
 142 IARPDLSLEPFDLSVKQTHVP--NLFSLQLCGAGFPINQSEVLASVGGSMIIGIDHSLY 200  
 186 IA--SSQATPVFDNMWQGLIPQDLFSVYLSOG---QSGSYVFGVDNSY 233  
 201 TGSIMWTPIRREWYEVIIIVRVEINGDQL--KMDKEYNYDKSIVDSGTTNLRPKKVF 258  
 234 SGLNWVPLTAETHYWOITVDSISIGGOVIACSGSC-----QAIVDTGTSLLAGP-- 286  
 259 AAVKASIKAAASSTEKFPDGFWMGEOLV--CWQAGTTPWNTFFVLSYLMGEVINOSEFRITIL 317  
 287 ANIQNYIGASQDSN-----GQVINCNNISNMPTVVF-----TINGVOY 321  
 318 PQOY--LRPVEDVATSQDCKY--FAISOSSTGT-----VMGAVIMEGFYVVDRAKRGIFA 371  
 322 GVQYPLSPASVVRNQOQSGSSGFQAMNLPNTNSGDLWILGDVFIQYFTVFDRAANNYVIA 381

Query Match 14.2%; Score 306; DB 2; Length 387;  
 Best Local Similarity 27.1%; Pred. No. 7.2e-18;  
 Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVG---AAPHFLHRYQRLSTYRDLRKGYYV 85  
 75 YFTGISIGTPPQSFVTFIDFGSSNLWVPSVYCSNPACTNHHMFNPQOSTFQATNTPVSI 134  
 86 PYTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAEI 142

Query Match 14.2%; Score 306; DB 2; Length 387;  
 Best Local Similarity 27.1%; Pred. No. 7.2e-18;  
 Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVG---AAPHFLHRYQRLSTYRDLRKGYYV 85  
 75 YFTGISIGTPPQSFVTFIDFGSSNLWVPSVYCSNPACTNHHMFNPQOSTFQATNTPVSI 134  
 86 PYTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAEI 142

J. Biochem. 129, 147-153, 2001  
 A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens  
 A:Reference number: JC7573; MUID:21064922; PMID:11134969  
 A:Contents: Stomach  
 A:Accession: JC7574  
 A:Molecule type: mRNA  
 A:Residues: 1-384 <IKU>  
 A:Cross-references: DDBJ:AB045380  
 A:Accession: PC7119  
 A:Molecule type: protein  
 A:Residues: 16-35; 57-76 <IK2>  
 A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a  
 C:Genetics:  
 A:Gene: Pga  
 C:Superfamily: pepsin  
 C:Keywords: stomach; zymogen

Query Match 14.3%; Score 308; DB 2; Length 384;  
 Best Local Similarity 27.5%; Pred. No. 4.8e-18;  
 Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL-----HRYQRLSTYRDLRKG 82  
 72 YGTGISIGTPPQSFVTFIDFGSSNLWV---PSVYCSQACSNHNFNPQOSTFQATNTP 128  
 83 YVPTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAE 141  
 129 VSIQYGTGSMGFLGYDTQV---GNIQISNQMFGLSESEPGSFLYSPFDGILGLAFPS 185  
 142 IARPDLSLEPFDLSVKQTHVP--NLFSLQLCGAGFPINQSEVLASVGGSMIIGIDHSLY 200  
 186 IA--SSQATPVFDNMWQGLIPQDLFSVYLSOG---QSGSYVFGVDNSY 233  
 201 TGSIMWTPIRREWYEVIIIVRVEINGDQL--KMDKEYNYDKSIVDSGTTNLRPKKVF 258  
 234 SGLNWVPLTAETHYWOITVDSISIGGOVIACSGSC-----QAIVDTGTSLLAGP-- 286  
 259 AAVKASIKAAASSTEKFPDGFWMGEOLV--CWQAGTTPWNTFFVLSYLMGEVINOSEFRITIL 317  
 287 ANIQNYIGASQDSN-----GQVINCNNISNMPTVVF-----TINGVOY 321  
 318 PQOY--LRPVEDVATSQDCKY--FAISOSSTGT-----VMGAVIMEGFYVVDRAKRGIFA 371  
 322 GVQYPLSPASVVRNQOQSGSSGFQAMNLPNTNSGDLWILGDVFIQYFTVFDRAANNYVIA 381

Query Match 14.3%; Score 308; DB 2; Length 384;  
 Best Local Similarity 27.5%; Pred. No. 4.8e-18;  
 Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL-----HRYQRLSTYRDLRKG 82  
 72 YGTGISIGTPPQSFVTFIDFGSSNLWV---PSVYCSQACSNHNFNPQOSTFQATNTP 128  
 83 YVPTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAE 141  
 129 VSIQYGTGSMGFLGYDTQV---GNIQISNQMFGLSESEPGSFLYSPFDGILGLAFPS 185  
 142 IARPDLSLEPFDLSVKQTHVP--NLFSLQLCGAGFPINQSEVLASVGGSMIIGIDHSLY 200  
 186 IA--SSQATPVFDNMWQGLIPQDLFSVYLSOG---QSGSYVFGVDNSY 233  
 201 TGSIMWTPIRREWYEVIIIVRVEINGDQL--KMDKEYNYDKSIVDSGTTNLRPKKVF 258  
 234 SGLNWVPLTAETHYWOITVDSISIGGOVIACSGSC-----QAIVDTGTSLLAGP-- 286  
 259 AAVKASIKAAASSTEKFPDGFWMGEOLV--CWQAGTTPWNTFFVLSYLMGEVINOSEFRITIL 317  
 287 ANIQNYIGASQDSN-----GQVINCNNISNMPTVVF-----TINGVOY 321  
 318 PQOY--LRPVEDVATSQDCKY--FAISOSSTGT-----VMGAVIMEGFYVVDRAKRGIFA 371  
 322 GVQYPLSPASVVRNQOQSGSSGFQAMNLPNTNSGDLWILGDVFIQYFTVFDRAANNYVIA 381

Query Match 14.3%; Score 308; DB 2; Length 384;  
 Best Local Similarity 27.5%; Pred. No. 4.8e-18;  
 Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL-----HRYQRLSTYRDLRKG 82  
 72 YGTGISIGTPPQSFVTFIDFGSSNLWV---PSVYCSQACSNHNFNPQOSTFQATNTP 128  
 83 YVPTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAE 141  
 129 VSIQYGTGSMGFLGYDTQV---GNIQISNQMFGLSESEPGSFLYSPFDGILGLAFPS 185  
 142 IARPDLSLEPFDLSVKQTHVP--NLFSLQLCGAGFPINQSEVLASVGGSMIIGIDHSLY 200  
 186 IA--SSQATPVFDNMWQGLIPQDLFSVYLSOG---QSGSYVFGVDNSY 233  
 201 TGSIMWTPIRREWYEVIIIVRVEINGDQL--KMDKEYNYDKSIVDSGTTNLRPKKVF 258  
 234 SGLNWVPLTAETHYWOITVDSISIGGOVIACSGSC-----QAIVDTGTSLLAGP-- 286  
 259 AAVKASIKAAASSTEKFPDGFWMGEOLV--CWQAGTTPWNTFFVLSYLMGEVINOSEFRITIL 317  
 287 ANIQNYIGASQDSN-----GQVINCNNISNMPTVVF-----TINGVOY 321  
 318 PQOY--LRPVEDVATSQDCKY--FAISOSSTGT-----VMGAVIMEGFYVVDRAKRGIFA 371  
 322 GVQYPLSPASVVRNQOQSGSSGFQAMNLPNTNSGDLWILGDVFIQYFTVFDRAANNYVIA 381

Query Match 14.3%; Score 308; DB 2; Length 384;  
 Best Local Similarity 27.5%; Pred. No. 4.8e-18;  
 Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;  
 30 YVEMTVGSPPTLNLIVDTGSSNFAVGAAPHFL-----HRYQRLSTYRDLRKG 82  
 72 YGTGISIGTPPQSFVTFIDFGSSNLWV---PSVYCSQACSNHNFNPQOSTFQATNTP 128  
 83 YVPTQKWEGLDLSIPHGPNVTYRANIAATESDK--FFINGSNWEGILGLAYAE 141  
 129 VSIQYGTGSMGFLGYDTQV---GNIQISNQMFGLSESEPGSFLYSPFDGILGLAFPS 185  
 142 IARPDLSLEPFDLSVKQTHVP--NLFSLQLCGAGFPINQSEVLASVGGSMIIGIDHSLY 200  
 186 IA--SSQATPVFDNMWQGLIPQDLFSVYLSOG---QSGSYVFGVDNSY 233  
 201 TGSIMWTPIRREWYEVIIIVRVEINGDQL--KMDKEYNYDKSIVDSGTTNLRPKKVF 258  
 234 SGLNWVPLTAETHYWOITVDSISIGGOVIACSGSC-----QAIVDTGTSLLAGP-- 286  
 259 AAVKASIKAAASSTEKFPDGFWMGEOLV--CWQAGTTPWNTFFVLSYLMGEVINOSEFRITIL 317  
 287 ANIQNYIGASQDSN-----GQVINCNNISNMPTVVF-----TINGVOY 321  
 318 PQOY--LRPVEDVATSQDCKY--FAISOSSTGT-----VMGAVIMEGFYVVDRAKRGIFA 371  
 322 GVQYPLSPASVVRNQOQSGSSGFQAMNLPNTNSGDLWILGDVFIQYFTVFDRAANNYVIA 381

```

Db      135 TYTGSMTGLGYDTVKV---GNIEDTNQIFGLSKTEPGITFLV---APFDGILGLAYPSI 189
Qy      143 ARPDDSLPEPFDLSLVKQTHV--PNLFSLQLCAGAGPPLNQSLVSLASVCGSMIIIGIDHSLYT 201
      :   :   |||:   :   :|||:|   :   :|||:|   :   :|||:|   :   :|||:|   :
Db      190 SASDAT--PVFDNMWNGELYSDELFSVYLSNG-----EKGSMVMFGGIDSSYYT 237
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy      202 GSLWYTFIREWYVEYIIVRVEINGODLKM--DCKEYNDKSIDVSGTTLNLRPKKVEA 259
      ||| :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      238 GSLNWVPVSHGHWQITMDSITINGETIACADSC-----QAVVDITGTSLLAGPPTSATSK 291
      ||| :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy      260 AVKSIKAASSTKFPDGFGLGEOLV--CWQAGTTPWNIPFVLSLYLMGEVYNOSFRITILP 318
      ||| :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      292 IQSVIGASKNL-----LGENIISCSAIDSLDPIV-----TINN 325
      ||| :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy      319 QQYLRPVED--VATSQDDC---YKFAISQSGTGT--VMGAVIMEGFVYVDFRARRKRGFAV 372
      ||| :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      326 VQPLPASAYILKEDDDCLSGFGDMMLDTSYGELWTLGDFVFIQYFTTFVDRANNOVGGLA 385
      ||| :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy      373 SA 374
      :
Db      386 AA 387

```

RESULT 5

pepsinogen C - African clawed frog  
JC7573

N:Alternate names: progastricsin  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7573; PC7118  
R:Ikuizawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.  
J. Biochem. 129, 147-153, 2001

A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens  
A:Reference number: JC7573; MUID:21064922; PMID:11134969  
A:Contents: Stomach  
A:Accession: JC7573  
A:Molecule type: mRNA  
A:Residues: 1-383 <IKU>  
A:Cross-references: DDBJ:AB045379  
A:Accession: PC7118  
A:Molecule type: protein  
A:Residues: 17-68 <IK2>  
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like activity.  
C:Genetics:  
A:Gene: Pgc  
A:Superfamily: pepsin  
C:Keywords: stomach; zymogen

[illegible]

```

QY   306  ---EVTNQ-----FRITILPQQYLRLRPVDTSQDDCYKFAISQSSTGTVMGAVIME    355
      |||         ||||       |:|          ::|||:
Db   326  PSAYVLQQSSGYCTIGIMPTLPSPNGQL-----WILDGVFLR    364
      :|::        |::|::     ::|||::|:
QY   356  GFVVDFDRARKRGIFAVSA    374
      :|:|        ::|||::|:
Db   365  EYSYVDLGNQVGAFATA    383
      :|:|        ::|||::|:

RESULT                6
SIPsin A (EC 3.4.23.1) 4 precursor - Japanese macaque
N:Alternate names: pepsinogen A isozyme 4
C:Species: Macaca fuscata (Japanese macaque)
C>Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
C:Accession: S19682; SI6065
R:kageyama, T.; Tanabe, K.; Koikwai, O.
Eur. J. Biochem. 202, 205-215, 1991
A>Title: Development-dependent expression of isozymogens of monkey pepsinogens and st
A:Reference number: S19681; MUID:92037645
A:Accession: S19682
A:Molecule type: mRNA
A:Residues: 1-368 <KAG>
A:Cross-references: EMBL:X59753; NID:g38070; PIDN:CAA42425.1; PID:g38071
A>Note: parts of sequence, including amino ends of pepsinogen and activation intermed
C:Comment: This is a minor component of pepsin at all post-partum stages.
C:Comment: Although two-step activation is observed, activation is predominantly a o
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-388/Product: pepsinogen A 4 #status experimental <PPT>
F:16-62/Domain: activation peptide #status experimental <APT>
F:63-388/Product: pepsin A 4 #status experimental <ENZ>
F:38-39/Cleavage site: Leu-Lys (pepsin) #status experimental
F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental
F:94,277/Active site: Asp #status predicted
F:107-112,268-272,311-344/bisulfide bonds: #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match                      14.0%; Score 302.5; DB 1; Length 388;
Best Local Similarity           27.6%; Pred. No.1.4e-17;
Matches                          107; Conservative               64; Mismatches 133; Indels 83; Gaps 17;
```

	Query Match	13.9%;	Score	299.5;	DB 2:	Length	396;	
	Best Local Similarity	25.9%;	Pred.	No. 2.6e-17;				
	Matches	100;	Conservative	65;	Mismatches	148;	Gaps	16;
QY	3	DEEPEEPGRGGSFVEMVDNLKRGSGGYVEMTVGPSPQTLLNLDVTGSSNFVCA	---	58				
		:     :	:	:    :	:	:		
Dd	63	DQSAAKEP-----LNYLD-----MEYFGTISGGPPQNFTVFDTGSSNLWFSVYCT	110					
		:	:	:    :	:	:		
QY	59	APHPFLHRYORQLSSSTYRDLRKGVYPYTQGKWGEELGTDLVSIPHGPNVIVRAIAAI	118					
		:     :	:	:    :	:	:		
Dd	111	SPACKTHSRFPQSOSSTYSPOGFSIQYGTGSLGIIGADQVSV--EGLTVVVGQGFGEV	169					
		:     :	:	:    :	:	:		
QY	119	TESDRFFINGSNWEGILGLVAETAREDDLSLEFFPSLVKQTHVPNLPLQLCGAGFPLN	178					
		:     :	:	:    :	:	:		
Dd	170	TPPGQTFVD-AFFDGILGLGPSIA---VGGVTPVFEDNNMAQ-----NLVLDPMFVSVMSSN	222					
		:     :	:	:    :	:	:		
QY	179	QSEVIASVGGSMITGGIDHSIYTGSLMYTPIRREYYEVTVIRVEINGDLMCKCEKYN	238					
		:     :	:	:    :	:	:		

Db 223 PE---GGAGSELIFGGYDHSFSGSLNWPVTKAYQIALDNTQVGG---TMFCSB--G 275  
Qy 239 DKSIVDSGTTNLRPKKVFEEAAVKSIAASTKFPDGFWMGEOLVQWQAGTTPWNIFPV 298  
Db 276 COAIVDTGTSITGSDKIKOLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322  
Qy 299 ISYLMGEVNTQSFRIITLPQOYLPRVEDVATSDCCYKFAISQSSTG----- 346  
Db 323 VTFTING-----VPYTLSPAY---TLDFVGMQFC-----SSGFGGLDIHPAGP 366  
Qy 347 -TVMGAVIMEGFYVDFRANKRIGFA 371  
Db 367 LWILGDVFIQFYSVFDRGNRRVGLA 392

RESULT 9  
A41443  
pepsin (EC 3.4.23.-) precursor, embryonic - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 21-Jul-2000  
C:Accession: A41443  
R:Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.  
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsin  
A:Reference number: A41443; MUID:88227903  
A:Accession: A41443  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-383 <HAY>  
A:Cross-references: GB:D00215; NID:g2760810; PIDN:BAA00153.1; PID:g222853  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 13.8%; Score 298.5; DB 2; Length 383;  
Best Local Similarity 25.2%; Pred. No. 3e-17;  
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;  
Qy 30 YVEMTVGSPPTNLILVDTGSSNFAVGA-----APHPFLHRYQRLSSYRDLRKGVYV 85  
Db 76 YGTISGTPPDFTVDFDTGSSNLWPSVCTSPACQSHQMFNPSSQSTYKSTGQNLIS 135  
Qy 86 PYTGCKWEGELGTDLVSIHPGPNVTVRANIAAITESSDKFFINGSNWEGILGLAYAEATRP 145  
Db 136 HYGTGDMETGVCDDVTVASLMDTNOLFGLST-SEPGGFVY-VKFDGILGLGYPSLAA- 192  
Qy 146 DDLSEPFDSLQKTHV-PNLFSLQLCAGFPLNQSEVLASVSGSMIIGGIDHSLYTGSL 204  
Db 193 -DGITPVFDNNVNSLLEQNLFVYLS-----REPMGSMVYFGGIDBSYFTGSI 240  
Qy 205 WYPIRREWYEVILVIRVEINGQDL--KMDCKEYNDKSIYDSTGTTNLRPKKVFEEAAVK 262  
Db 241 NWIPVSGYQWQISMDSIYNKQBIACSSG-----QALIDGTSLVAGPASDINDIQS 294  
Qy 263 SIKAASTEKPPDGFWMGEOLVQWQAGTTPWNIFPVISL-----YLMGEVNTQSFRIITLP 318  
Db 295 AVGANQNT-----YGEYSV-----NCSHILAMDVVVFVIGGI----- 326  
Qy 319 QOYLPRVEDVA-----TSQDCCYKFAISQSTGTVMGAVIMEGFYVDFRANKRIGFA 371  
Db 327 -QY--PVPALAYTEQCGQCTMSSFQNSADLWILGDVFIQFYSVFDRANNRVGLA 380

RESULT 10  
A39314  
gastricsin (EC 3.4.23.3) precursor - bullfrog  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 19-Jun-1992 #sequence\_revision 19-Jun-1992 #text\_change 22-Jun-1999  
C:Accession: A39314  
R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageya  
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and pep

A:Reference number: A39314; MUID:92042186  
A:Accession: A39314

A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-384 <YAK>  
A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 13.8%; Score 298.5; DB 2; Length 384;  
Best Local Similarity 25.5%; Pred. No. 3e-17;  
Matches 100; Conservative 61; Mismatches 142; Indels 89; Gaps 15;  
Qy 14 SFVEMVDNLCKSGGQYVEMTVGSPPTNLILVDTGSSNFAVGAAPHPFL-----HR 66  
Db 51 NFATAFEPLANMDSYGEISIGTPPQNLVLFDTGSSNLWV---PSTYCSQSQACTNHP 107  
Qy 67 YYORQLSSTYRDLRKGVYVPTQCKWEGELGTDLVSIHPGPNVTVRANIA-----AI 118  
Db 108 QFNPSQSSSYSSNQOQFSLQVGTGSLGILGYDVIQI-----QNTAISQOEFGLSV 158  
Qy 119 TESDKFFINGSNWEGILGLAYAEATRPDSDLEPFDFSLVKQTHVPN-LFSLQLCGAGFPL 177  
Db 159 TEPGTNFVY-AQFDGILGLAYPSIA--EGGATTVMQMIQOONLINOPLFAFYLSGQONSQ 215  
Qy 178 NQSEVLASVSGSMIIGGIDHSLYTGSLWYPIRREWYEVILVIRVEINGQD---LKMDC 234  
Db 216 N-----GGEAVFGVDQNYVSGQIYWTVPVSETYMQIGQGFVNGOATGWCSCGC- 266  
Qy 235 EYNVDKSIYDSTGTTNLRPKKVFEEAAVKSIAASTKFPDGFWMGEOLV-CWQAGTTPW 293  
Db 267 -----QGIVDTGTSLLTAPQSVEFSLMQSIGAQDQN-----GOYAVSCSNISQSLPT 313  
Qy 294 NIFPVI-----SLYLMGEVNTQ---SFRITLPQOYLPRVEDVATSDCCYKFAISQ 342  
Db 314 ISFTISGVSFPLPPSAVYVLAQNSGYCTIGIMPTVLPYLSQNGQPL----- 356  
Qy 343 SSTGTWAGVIMEGFYVDFRANKRIGFAVSA 374  
Db 357 ----WILGDVFLRQYYSVYDLGNQNVGFAAAA 384

RESULT 11  
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 23-Feb-1997  
C:Accession: C38302  
R:Kageyama, T.; Tanabe, K.; Koiwai, O.  
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu  
A:Reference number: A38302; MUID:91009127  
A:Accession: C38302  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 13.7%; Score 296; DB 2; Length 387;  
Best Local Similarity 26.9%; Pred. No. 5e-17;  
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;  
Qy 30 YVEMTVGSPPTNLILVDTGSSNFAVGAAPHPF-----LHRYQRLSSYRDLRK 82  
Db 75 YFGTISGTPPDFTVDFDTGSSNLWV---PSTYCSLALHKNRPEDSDSYQGTSET 131  
Qy 83 VYVPYTGCKWEGELGTDLVSIHPGPNVTVRANIAAITESSDKFFINGSNWEGILGLAYAEI 142  
Db 132 LSITYGTGSMTGILGYDVIQVKSIEDTNQIFGLSKTEPSLTFLF--APFDGILGLAYPSI 189





Wed Aug 7 10:38:05 2002

QY	177	LNQSEVLASVGGSMIIIGGDHSLYTGTSLWYTPTRREYYEVIIVRVEINGDQLKMDCKEY	236
		:    :    :    :  :	:
Db	221	-DQS-----GSAVFIRGDSYYTGGLNNWPSPVEGYWQLSVDSITNNGEAIA--CAE-	270
QY	237	NYDKSIVDSGTTNLRLPKPVFAA VKSIKAASSTEFDPDGFLWGQLV-CWQAQTTPWNI	295
		:      :    :  :	:
Db	271	-GCQAIVDVTGSLLTGTPTSPIANQSDICASENSD-----GEWVVSCEISSLPDIV	321
QY	296	FPVISILMGEVTNQSFRTIIPQQYLRPVEDVATSQDDCYK-----FAISOSSTGTVMWG	350
		:  :  :	:
Db	322	F-----TINGIQVPVPBSAY-----ILQSGSCISGFQGMDPVPTESGELWI LG	364
QY	351	AVIMEGYPVVVDFDRARKRGFA	371
		:  :  :    :  :	:
Db	365	DVFIRQYFTVTFDRANNQVGILA	385

Search completed: August 7, 2002, 09:14:18  
Job time: 42 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:15:11 ; Search time 17.62 Seconds  
(without alignments)  
894.373 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 2156

Sequence: 1 ETDEPEPGRGSRFVEMV.....GPFVTLDMEDCGYNIPQTDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2156	100.0	501	1	BACE_HUMAN
2	2134	99.0	501	1	BACE_RAT
3	2132	98.9	501	1	BACE_MOUSE
4	1132	52.5	518	1	BACE2_HUMAN
5	327	15.2	324	1	PEP1_GADMO
6	314.5	14.6	390	1	CATD_BOVIN
7	306	14.2	387	1	PEP1_RABIT
8	302.5	14.0	388	1	PEP4_MACFU
9	302	14.0	367	1	CATE_HUMAN
10	299.5	13.9	396	1	PEPE_CHICK
11	298.5	13.8	383	1	PEPE2_RABIT
12	296	13.7	387	1	CATE_CAYPO
13	296	13.7	391	1	CATD_HUMAN
14	295.5	13.7	412	1	PEP4_RABIT
15	295	13.7	387	1	PEP4_RABIT
16	289.5	13.4	388	1	PEP2_MACFU
17	287	13.3	407	1	CATD_RAT
18	286	13.3	387	1	PEP3_RABIT
19	285.5	13.2	398	1	CATE_RAT
20	285	13.2	398	1	CATD_CHICK
21	283.5	13.1	388	1	PEPA_MACMU
22	282.5	13.1	388	1	PEPA_HUMAN
23	281.5	13.1	381	1	CHYM_SHEEP
24	281.5	13.1	388	1	PEP1_MACFU
25	279.5	13.0	410	1	CATD_MOUSE
26	278.5	12.9	386	1	PEPA_PIG
27	278	12.9	387	1	PEPA_CALJA
28	277.5	12.9	396	1	CATE_RABIT
29	277.5	12.9	397	1	CATE_MOUSE
30	276.5	12.8	419	1	CARV_CANAL
31	273.5	12.7	381	1	CHYM_BOVIN
32	273	12.7	388	1	PEP1_RABIT
33	270.5	12.5	377	1	PEP2_MACFU

34	270	12.5	376	1	PAG2_BOVIN	Q28057 bos taurus
35	267.5	12.4	381	1	CHYM_CALJA	Q9n2d2 callithrix
36	267	12.4	365	1	CATD_SHEEP	Q9mzs8 ovis aries
37	266	12.3	405	1	CARP_YEAST	P07267 saccharomyc
38	264	12.2	396	1	CARP_NEUCR	Q01294 neurospora
39	263.5	12.2	388	1	PEPC_CALJA	Q9n2d3 callithrix
40	262.5	12.2	394	1	PEPC_CAYPO	Q644l1 cavia porce
41	261.5	12.1	387	1	ASPP_AEIDAE	Q03168 aedes aegyp
42	261.5	12.1	388	1	PEPC_HUMAN	P20142 homo sapien
43	261	12.1	388	1	PAG_HORSE	Q28389 equus cabal
44	260	12.1	345	1	CATD_PIG	P00795 sus scrofa
45	255	11.8	496	1	ASPR_ORYSA	P42211 oryza sativ

## ALIGNMENTS

RESULT 1  
BACE\_HUMAN  
ID BACE\_HUMAN STANDARD; PRT; 501 AA.  
AC P56817; Q9UJT5;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)  
DE (Memapsin-2).  
GN BACE OR BACE1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RC TISSUE=Brain;  
RX MEDLINE=20002972; PubMed=10531052;  
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
RA Treanor J., Rogers G., Citron M.;  
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
the transmembrane aspartic protease BACE.";  
RL Science 286:735-741(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND  
RC CHARACTERIZATION.  
RX TISSUE=Brain;  
RA Sinha S., Anderson J.P., Barbour R., Basl G.S., Caccavello R.,  
RA Davis D., Doan M., Dovey H.F., Frigon N., Hong J., Jacobson-Crosk K.,  
RA Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,  
RA Tatsuno G., Tung J., Schenk D., Seubert P., Suenensaar S.M., Wang S.,  
RA Walker D., Zhao J., McConlogue L., Varghese J.;  
RT "Purification and cloning of amyloid precursor protein beta-secretase  
from human brain".  
RL Nature 402:537-540(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=20057170; PubMed=10591213;  
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,  
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;  
RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-  
secretase activity".  
RL Nature 402:533-537(1999).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=20030166; PubMed=10561122;  
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,  
RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,  
RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;

"Identification of a novel aspartic proteinase (ASP 2) as beta-secretase.";  
Mol. Cell. Neurosci. 14:419-427(1999).  
[5]  
SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE=Brain, and Pancreas;  
RA Michel B., De Pietro Tonelli D., Zaccchetti D., Keller P.;  
RT "New beta-site APP cleaving enzyme isoform (BACE-1B) obtained from  
human brain and pancreas.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
[6]  
SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.  
RX MEDLINE-20144060; PubMed-10677483;  
RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;  
RT "Human aspartic protease menapsin 2 cleaves the beta-secretase site of  
beta-amyloid precursor protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).  
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC B/BACE-1B; are produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: BRAIN.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AF190725; AAF04142.1; -  
CC DR EMBL; AF201466; AAF18982.1; -  
CC DR EMBL; AF200343; AAF17079.1; -  
CC DR EMBL; AF204943; AAF26367.1; -  
CC DR EMBL; AF338816; AAK38374.1; -  
CC DR EMBL; AF200193; AAF13715.1; -  
CC DR HSSP; P32329; 1YPS.  
CC MIM; 604252; -  
CC DR InterPro; IPR001969; Asp.protease.  
CC DR InterPro; IPR001461; Pepsin.  
CC DR Pfam; PF00026; asp; 3.  
CC DR PRINTS; PR00792; PEPSIN.  
CC DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
CC KW hydrolase; Aspartyl protease; Glycoprotein; zymogen; Transmembrane;  
CC FT Signal; Alternative splicing. POTENTIAL.  
CC FT SIGNAL; 1 21  
CC FT PROPEP; 22 45  
CC FT CHAIN; 46 501  
CC FT DOMAIN; 22 457  
CC FT TRANSMEM; 458 478  
CC FT DOMAIN; 479 501  
CC FT ACT\_SITE; 93 93  
CC FT ACT\_SITE; 289 289  
CC FT CARBOHYD; 153 153  
CC FT CARBOHYD; 172 172  
CC FT CARBOHYD; 223 223  
CC FT CARBOHYD; 354 354  
CC FT CARBOHYD; 354 354  
CC FT VARSPIC; 190 214  
CC SEQUENCE 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;  
Query Match 100.0%; Score 2156; DB 1; Length 501;  
Best Local Similarity 100.0%; Pred. No. 1.3e-171;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPGRGSGFVEMVNDLRKSGQGYVEMTVGSPPTLNILNLDVTGSSNFAVGAAP 60  
DB 46 ETDEPEPGRGSGFVEMVNDLRKSGQGYVEMTVGSPPTLNILNLDVTGSSNFAVGAAP 105  
QY 61 HPFLHRYQROLSTYRDLRKGVVVPYTOGKWEGLGTLVSIHPGPNVTVRANIAITE 120  
DB 106 HPFLHRYQROLSTYRDLRKGVVVPYTOGKWEGLGTLVSIHPGPNVTVRANIAITE 165  
QY 121 SDKFFINGSNWEGLGLAYAEIARPDSDLEFPFDSLKYQTHVPLNLSIQLCGAGFPLNQS 180  
DB 166 SDKFFINGSNWEGLGLAYAEIARPDSDLEFPFDSLKYQTHVPLNLSIQLCGAGFPLNQS 225  
QY 181 EVLASVGSMTIGGDHSLYTGSLWYTPIRREWYEVIIIRVEINGQDLKMDCKEYNDK 240  
DB 226 EVLASVGSMTIGGDHSLYTGSLWYTPIRREWYEVIIIRVEINGQDLKMDCKEYNDK 285  
QY 241 SIIVDSGTTNLRPKKVFEEAAVKSIAKASTKFPDGFGLGOLVCWQAGTTPWNIFFPVS 300  
DB 286 SIIVDSGTTNLRPKKVFEEAAVKSIAKASTKFPDGFGLGOLVCWQAGTTPWNIFFPVS 345  
QY 301 LYLMGEVNTQSRITLPOQYLRPVEDVATSDDCYKFAISQSSGTVMGAVIMEGFYV 360  
DB 346 LYLMGEVNTQSRITLPOQYLRPVEDVATSDDCYKFAISQSSGTVMGAVIMEGFYV 405  
QY 361 FDRARRKRGFAYSACHVHDEFRTAAVEGPFVTLDMDCGYNIPQTD 407  
DB 406 FDRARRKRGFAYSACHVHDEFRTAAVEGPFVTLDMDCGYNIPQTD 452  
RESULT 2  
BACE-RAT STANDARD; PRT; 501 AA.  
ID BACE-RAT  
AC P56819;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Beta-secretase precursor (EC 3.4.23.-) (beta-site APP cleaving enzyme)  
DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
DE protease 2) (ASP 2) (ASP2) (Membrane-associated aspartic protease 2)  
DE (Memapsin-2).  
GN BACE.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20002972; PubMed=10531052;  
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
RA Fisher S., Fuller J., Edenson S., Lille J., Jarosinski M.A.,  
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
RA Treanor J., Rogers G., Citron M.;  
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
the transmembrane aspartic protease BACE.";  
RL Science 286:735-741(1999).  
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

```

or send an email to license@isb-sib.ch).
-----
CC EMBL; AF190727; AAF04144.1; -.
CC HSSP; P32329; 1YPS.
CC DR InterPro; IPR001969; Asp_protease.
CC DR Pfam; PF00026; asp; 3.
CC DR PRINTS; PR00792; PEPSIN.
CC DR PROSITE; PS00141; ASP_PROTEASE; 1.
CC KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
CC FT SIGNAL 1 21 POTENTIAL.
CC FT PROPEP 22 45 POTENTIAL.
CC FT CHAIN 46 501 BETA-SECRETASE.
CC FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 458 478 POTENTIAL.
CC FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
CC FT ACT_SITE 93 93 BY SIMILARITY.
CC FT ACT_SITE 289 289 BY SIMILARITY.
CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 501 AA; 55806 MW; 24B445BCB8E7DE3 CRC64;

Query Match 99.0%; Score 2134; DB 1; Length 501;
Best Local Similarity 98.3%; Pred. No. 8.7e-170;
Matches 400; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ETDPEEPGRGSGFVEMVDNLGRKSGGQYVMTVGGSPPTNLILVDTGSSNFAVGAAP 60
DB 46 ETDPEEPGRGSGFVEMVDNLGRKSGGQYVMTVGGSPPTNLILVDTGSSNFAVGAAP 105

QY 61 HPFLHRYQRLSSYRDLKRGVYVYTGKGWEGELGDLVSLIPGPNVYRANIAATE 120
DB 106 HPFLHRYQRLSSYRDLKRGVYVYTGKGWEGELGDLVSLIPGPNVYRANIAATE 165

QY 121 SDKFFINSNWEGLGLAYAEIARDDSLFFDLSLVKQTHVPLNLSLQCGAGFPNLQS 180
DB 166 SDKFFINSNWEGLGLAYAEIARDDSLFFDLSLVKQTHVPLNLSLQCGAGFPNLQT 225

QY 181 EYLASVGGSMIIGDHSLYTGLSWYTPIRREWYEVILIVRVEINGDLKMDCKEYNDK 240
DB 236 EYLASVGGSMIIGDHSLYTGLSWYTPIRREWYEVILIVRVEINGDLKMDCKEYNDK 285

QY 241 STVDSGTTNLRPLPKVFAAVKSIKAASSTKFPDGFGLGQLVCWQAGTTPWNIFFVIS 300
DB 286 STVDSGTTNLRPLPKVFAAVKSIKAASSTKFPDGFGLGQLVCWQAGTTPWNIFFVIS 345

QY 301 LYLMEVTNQSFRTILPQQYLRLPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYV 360
DB 346 LYLMEVTNQSFRTILPQQYLRLPVEDVATSDQDCYKFAISQSTGTVMGAVIMEGFYV 405

QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMDCGYNIPOTDE 407
DB 406 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMDCGYNIPOTDE 452

RESULT 3
BACE_MOUSE
ID BACE_MOUSE STANDARD; PRT; 501 AA.
AC P56818;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)
DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl
DE protease 2) (ASP 2) (ASP2) (Membrane-associated aspartic protease 2)
DE (Memapsin-2).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lille J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE.";
RL Science 286:735-741(1999).
[2]
RP REVISIONS TO 6 AND 81-87.
RA Bennett B.D., Vassar R., Citron M.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20051710; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
RT "Membrane-anchored aspartyl protease with Alzheimer's disease
RT beta-secretase activity.";
RL Nature 402:533-537(1999).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
CC THE A-BETA PEPTIDE SEQUENCE. BETWEEN RESIDUES 671 AND 672 OF APP,
CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF190726; AAF04143.2; -.
CC EMBL; AF200346; AAF17082.1; -.
CC HSSP; P56272; 1AM5.
CC MGD; MGI:1346542; Bace.
CC InterPro; IPR001969; Asp_protease.
CC InterPro; IPR001461; Pepsin.
CC Pfam; PF00026; asp; 1.
CC PRINTS; PR00792; PEPSIN.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
CC KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
CC FT SIGNAL 1 21 POTENTIAL.
CC FT PROPEP 22 45 POTENTIAL.
CC FT CHAIN 46 501 BETA-SECRETASE.
CC FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 458 478 POTENTIAL.
CC FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
CC FT ACT_SITE 93 93 BY SIMILARITY.
CC FT ACT_SITE 289 289 BY SIMILARITY.
CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 501 AA; 55747 MW; C085A013145E474E CRC64;

```

Query Match

98.9%; Score 2132; DB 1; Length 501;

Best Local Similarity 98.3%; Pred. No. 1.3e-169;		Matches 400; Conservative 4; Mismatches 3; Indels 0; Gaps 0;	
Qy	1	ETDEPEPEGRGSGFVEMVNLKRGKSGQYVEMTVGSPPTLNILVDTGSSNFVGAAP	60
Db	46	ETDESEEPGRGSGFVEMVNLKRGKSGQYVEMTVGSPPTLNILVDTGSSNFVGAAP	105
Qy	61	HPFLHRYQRLSTYSYRDLKRGVVPYVTOGKWEGLGTDLVSPHGNVTVRANIAITE	120
Db	106	HPFLHRYQRLSTYSYRDLKRGVVPYVTOGKWEGLGTDLVSPHGNVTVRANIAITE	165
Qy	121	SDKFFINGSNWEGLGLAYAEIARPDSDLEPFDSLVKQTHVNLFLSLQLCGAGFPLNQS	180
Db	166	SDKFFINGSNWEGLGLAYAEIARPDSDLEPFDSLVKQTHVNLFLSLQLCGAGFPLNQT	225
Qy	181	EVLASVGSMIIGGIDHSLYTGSLSWYTPIRREWYVEIIVRVEINGDLKMDCKEYNYDK	240
Db	226	EALASVGSMIIGGIDHSLYTGSLSWYTPIRREWYVEIIVRVEINGDLKMDCKEYNYDK	285
Qy	241	SIVDSGTTNLRPKKVFVFAAVKSKAASSTKFPDGFGLGQVLCWQAGTTPWNIFFVIS	300
Db	286	SIVDSGTTNLRPKKVFVFAAVKSKAASSTKFPDGFGLGQVLCWQAGTTPWNIFFVIS	345
Qy	301	LYLMEVNTQSFRTILPQQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFYVV	360
Db	346	LYLMEVNTQSFRTILPQQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFYVV	405
Qy	361	FDRARKRIGFAVSACHVHDEFRTAAVEGFFVTLMDGCGYNIPTQDE	407
Db	406	FDRARKRIGFAVSACHVHDEFRTAAVEGFFVTLMDGCGYNIPTQDE	452
RESULT 4			
ID	BAE2_HUMAN	STANDARD;	PRT; 518 AA.
AC	Q9V5Z0; Q9UJTG;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving		
DE	enzyme 2) (Aspartyl protease 1) (Asp 1) (ASPI) (Membrane-associated		
DE	aspartic protease 1) (Memapsin-1).		
GN	BACE2 OR ASP21.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=20051710; PubMed=10591213;		
RX	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,		
RA	Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,		
RA	Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;		
RT	"Membrane-anchored aspartyl protease with Alzheimer's disease		
RT	beta-secretase activity.";		
RT	Nature 402:533-537(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Bone marrow;		
RA	Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,		
RA	Giese K.;		
RT	"Identification of a novel aspartic-like protease differentially		
RT	expressed in human breast cancer cell lines.";		
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Accarino M.P., Funagalli P., Ottolenghi S., Taramelli R.;		
RT	"Cloning of a gene from chromosome 21 Down region encoding a potential		
RT	transmembrane aspartyl protease.";		
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RA	Solans A., Estivill X., de la Luna S.;		

RT	"Cloning of a novel mammalian aspartyl protease.";	
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.	
RN	[5]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20030166; PubMed=10561122;	
RA	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,	
RA	Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,	
RA	Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;	
RT	"Identification of a novel aspartic proteinase (Asp 2) as	
RT	beta-secretase.";	
RL	Mol. Cell. Neurosci. 14:419-427(1999).	
RN	[6]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20144060; PubMed=10677483;	
RA	Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;	
RT	"Human aspartic protease memapsin 2 cleaves the beta-secretase site of	
RT	beta-amyloid precursor protein.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).	
RN	[7]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20289799; PubMed=10830953;	
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,	
RA	Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,	
RA	Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,	
RA	Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,	
RA	Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,	
RA	Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Antonarakis S.E.,	
RA	Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Brandt P.,	
RA	Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Kauer G., Bloeker H.,	
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Rieselmann L., Dagand E.,	
RA	Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,	
RA	Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,	
RA	Lehrach H., Reinhardt R., Yaspo M.-L.;	
RT	"The DNA sequence of human chromosome 21.";	
RL	Nature 405:311-319(2000).	
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.	
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE	
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; AF200342; AAF17078.1; -	
DR	EMBL; AF117892; AAD45240.1; -	
DR	EMBL; AF050171; AAD45963.1; -	
DR	EMBL; AF178532; AAF29494.1; -	
DR	EMBL; AF204944; AAF26368.1; -	
DR	EMBL; AF200192; AAF13714.1; -	
DR	EMBL; AL163284; CAB90458.1; -	
DR	EMBL; AL163285; CAB90554.1; -	
DR	HSP; P00797; 2REN.	
DR	MM; 605668; -	
DR	InterPro; IPR001969; Asp-protease.	
DR	InterPro; IPR001461; Pepsin.	
DR	Pfam; PF00026; asp; 3.	
DR	PRINTS; PR00792; PEPSIN.	
DR	PROSITE; PS00141; ASP_PROTEASE; 2.	
KW	Hydrolase; Aspartyl protease; Zymogen; Transmembrane;	
KW	Signal.	
FT	SIGNAL	1 20
FT	PROPEP	21 ?
FT	CHAIN	? 518
FT	DOMAIN	21 473
FT	TRANSMEM	474 494
FT	DOMAIN	495 518
FT	ACT_SITE	110 110
FT	ACT_SITE	303 303
FT	CARBOHYD	170 170
FT	POTENTIAL.	
FT	BETA SECRETASE 2.	
FT	EXTRACELLULAR (POTENTIAL).	
FT	POTENTIAL.	
FT	CYTOPLASMIC (POTENTIAL).	
FT	BY SIMILARITY.	
FT	BY SIMILARITY.	
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).	

FT CARBOHYD 366 366 N-LINKED (GLCNAC... (POTENTIAL).  
FT CONFLICT 36 36 A -> T (IN REF. 6).  
SQ SEQUENCE 518 AA; 56180 MW; 2E903150823760D3 CRC64;

Query Match 52.5%; Score 1132; DB 1; Length 518;  
Best Local Similarity 54.0%; Pred. No. 1.7e-86;  
Matches 209; Conservative 65; Mismatches 109; Indels 4; Gaps 2;

QY 9 PGRGSEFVMDNLGRKSGGYVEMTVGSPQTLNLLVDTGSSNFAVGAAPHPFLHRY 68  
Db 71 PAGAANFLAWNDLQGGSGRYLEMLIGTPQKQLILVDTGSSNFAVGAAPHPHSDYDTF 130  
QY 69 QRLSSYRDLRKGYYVPTQGGKEGELGDLVSIHPGNVTVRANIAAITEDEKKEFNG 128  
Db 131 DTERSSTYRSGFDVTKYTGQSWTGVGEDLVTPKGFNTSLVNIATTFESENFLPG 190  
QY 129 SNWEGILGLAYAEARTARDDSLRPFDSLVKQTHVNPFLSLQCGAGFPLNQSEVLASVGG 188  
Db 191 IKWNGILGLAYATLAKSSLETFDFSLVLTQANIPNVFSMQMGAGLPAVGS---GTNGG 247  
QY 189 SMIGGDHSLYGTSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTT 248  
Db 248 SLVGGTEPSLYKGDITWYPIKEWYQIETLKELEGGGSLNDDCREYNADKAIVDSGTT 307  
QY 249 NLRLPKVFEAAVKSIAASSTKFFDGFGLGQVLCVQAGTTPWNIPVISLYLMGEVNT 308  
Db 308 LLRLPKVFEAAVKSIAASSTKFFDGFGLGQVLCVQAGTTPWNIPVISLYLMGEVNT 308  
QY 309 NQSFRTILPOLYQPMGAGLNY-ECYRFGISPSNVALIGATVMEGVYVIFDRAQKRV 426  
Db 368 SRSFRTILPOLYQPMGAGLNY-ECYRFGISPSNVALIGATVMEGVYVIFDRAQKRV 426  
QY 369 GFAVSACHVDEPRTAAVEGPFVTLDM 395  
Db 427 GFAASPCAETAGAAVSEISGPFSTEDV 453

RESULT 5  
PEPL\_GADMO  
ID PEPL\_GADMO STANDARD; PRT; 324 AA.  
AC P56272;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Pepsin IIB (EC 3.4.23.-)  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;  
OC Gadus.  
OX NCBI\_TaxID=8049;  
RN [1]  
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.  
RC TISSUE=Stomach;  
RA "Structure and proposed amino-acid sequence of a pepsin from Atlantic cod (Gadus morhua).";  
RT cod (Gadus morhua).";  
RL Acta Crystallogr. D 54:32-46(1998).  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR PDB; 1AM5; 24-DEC-97.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; Asp\_PROTEASE; 2.  
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW 3D-structure.  
FT ACT SITE 32 32 BY SIMILARITY.  
FT ACT SITE 214 214 BY SIMILARITY.  
FT DISULFID 45 50 BY SIMILARITY.  
FT DISULFID 206 209 BY SIMILARITY.

FT DISULFID 247 280 BY SIMILARITY.  
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 15.2%; Score 327; DB 1; Length 324;  
Best Local Similarity 27.9%; Pred. No. 7.3e-20;  
Matches 104; Conservative 66; Mismatches 137; Indels 66; Gaps 15;

QY 18 MVDNLRKSGQGYVEMTVGSPQTLNLLVDTGSSNFAVGAAPHPFLHRYTORQLS 73  
Db 2 VTRQMKNEADTEYGVISITGTTPESKVFIEDTGSSMLWVSSSHCSQAQCSNNHKKFPROS 61  
QY 74 STYRDLRKGYVPTQGGKEGELGDLVSIHPGNVTVRANIAAITEDEKKEFNG 131  
Db 62 STYVETGKTVDLTGTGGMRLGQDTSVGGSDPNQELG---ESQTEPGFPQA-AAFP 117  
QY 132 EGLILGLAYAEARTARDDSLRPFDSLVKQTHVNPFLSLQCGAGFPLNQSEVLASVGGSM 190  
Db 118 DGLILGLAYAEARTARDDSLRPFDSLVKQTHVNPFLSLQCGAGFPLNQSEVLASVGGSM 190  
QY 191 IIGGDHSLYGTSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTN 249  
Db 167 -LGGVDNSHYTGSIIHWIPVTAEKYMQVALDGTITVNGQTAACEG-----QAIVDTGTSK 219  
QY 250 LRLPKVFEAAVKSIAASSTKFFDGFGLGQVLCVQAGTTPWNIPVISLYLMGEVNT 309  
Db 220 IVAPVSALANIMKDIGASEN-----QGENMGN---CASVQSLPDITF-----TI 260  
QY 310 QSFRTILPOLYQPMGAGLNY-ECYRFGISPSNVALIGATVMEGVYVIFDRAQKRV 426  
Db 261 NGVKQPLPSPAYIEGDAQFCTS-----GLGSGVPSNTSELWIFDGLFLRNYYTII 311  
QY 362 DRARKRIGFAVSA 374  
Db 312 DRTNKNKVGFAAA 324

RESULT 6  
CATD\_BOVIN  
ID CATD\_BOVIN STANDARD; PRT; 390 AA.  
AC P80209; Q9TS27;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cathepsin D precursor (EC 3.4.23.5).  
GN CTSD.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE OF 1-48.  
RC TISSUE=Milk;  
RA MEDLINE=93202276; PubMed=8454061;  
RA Larsen L.B., Boisen A., Petersen T.E.;  
RT "Procathepsin D cannot autoactivate to cathepsin D at acid pH.";  
RL FEBS Lett. 319:54-58(1993).  
RN [2]  
RP SEQUENCE OF 45-390, AND X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).  
RC TISSUE=Liver;  
RA MEDLINE=93223670; PubMed=8467789;  
RA Metcalf P., Fusek M.;  
RT "Two crystal structures for cathepsin D: the lysosomal targeting signal and active site.";  
RL EMBO J. 12:1293-1302(1993).  
CC -I- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR  
CC PROTEIN BREAKDOWN.  
CC -I- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,  
CC that of pepsin A. Does not cleave the 4-Gln-|-His-5 bond in B  
CC chain of insulin.  
CC -I- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.  
CC -I- SUBCELLULAR LOCATION: Lysosomal.

```

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR: S32383; S32383.
DR PIR: S37419; S37419.
DR HSP: P07339; 11VB.
DR MEROPS: A01.009; .
DR InterPro: IPR001969; Asp_protease.
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolase: Aspartyl protease; Glycoprotein; Lysosome; Zymogen.
FT PROPEP 45 390
FT CHAIN 77 77
FT ACT_SITE 273 273
FT ACT_SITE 71 140
FT DISULFID 90 97
FT DISULFID 264 268
FT DISULFID 307 344
FT CARBOHYD 114 114
FT CARBOHYD 241 241
SQ SEQUENCE 390 AA; 42488 MW; 5B38AA1C33C48D35 CRC64;

Query Match 14.6%; Score 314.5; DB 1; Length 390;
Best Local Similarity 28.0%; Pred. No. 1e-18;
Matches 113; Conservative 72; Mismatches 128; Indels 91; Gaps 21;

QY 8 EPG-RRGSFVEMVDNLKSGQGYVEMTGGSPPTLNLVDGTSSNFAVGAAPFL-- 64
DB 39 EPARVGGPIPELLKNYMDAQ---YIGEIGTTPQCFTVFDGTSANLWVPSIHCKLLDI 95
QY 65 ----HRYORQLSSYTRDLRKGVY-VPTQCKWEGELGTLVSIHPGN-----VTVR 112
DB 96 ACWTHRKYSKSSY--VKNGTTFDIHYGSGSLSGYLSQDTSVFCNPSSSPGCVIVQ 153
QY 113 ANI--AAITESDKFFNGWEGILGLAYAEIARPDLSLEPFDLSLVKQTHV-PNLFSIQ 169
DB 154 RQTGFAIKPGVWFI-AAKFDGILGMAYPRIS--VNNVLPVFDNLQKLVKDNVFS-- 208
QY 170 LCGAGFPLNOSVLAAGSGMIIGDHSYTGSLWYTPIREWYEVIVRVEINGQDL 229
DB 209 -----FELNR-DPKAQGGELMLGGTDSKYRGSLMFHNTVQAVQWQHMDQLDV-GSSL 261
QY 230 KMDCKEYNDKSIIVDSTTNLRPKVFEAAVKAASSTKFPDGFVLGEOLV-CWQA 288
DB 262 TV-CK--GGCEALVDGTSLVGPVEVRELQKAGVPLIQ-----GEVMIPEKV 310
QY 289 GTTPWNIFPVISLYLMGEVNTQSFRTIILPQQVLRVEDVATSDQDCYKFAISQSTGT- 347
DB 311 SS-----LPEVTVKLG-----KDYALSPED-YALKVSAQETTV 344
QY 348 -----VMGAVIMEGFYVVDVDRKRIGFAVSA 374
DB 345 LSGFMGMGDIPIPPGGLMILGDVFTIGRYITVFDQNRVGLAEAA 388

RESULT 7
PEP1_RABIT STANDARD; PRT; 387 AA.
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin II-1 precursor (EC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;

```

```

RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR: B38302; B38302.
DR HSP: P00791; 1PSA.
DR MEROPS: A01.001; .
DR InterPro: IPR001969; Asp_protease.
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolase: Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59
FT CHAIN 60 387
FT MOD_RES 129 129
FT ACT_SITE 93 93
FT ACT_SITE 276 276
FT DISULFID 106 111
FT DISULFID 267 271
FT DISULFID 310 343
FT SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;

Query Match 14.2%; Score 306; DB 1; Length 387;
Best Local Similarity 27.1%; Pred. No. 5.2e-18;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;

QY 30 YVEMTVGSPPTLNLVDGTSSNFAVG-----AAPHPLHRYQQLSTYDLRKGVYV 85
DB 75 YFGTISIGTPQEFVIEDTSSNLWVPSYCSLACFLHRRFDDSDSTFQATSETLSI 134
QY 86 PTQCKWEGELGTLVSIHPGNVTVRANIAAITESD---KFFINGSNWEGILGLAYAEI 142
DB 135 TYGTGSMTGILGYDTVKV---GNIEDTQIFGLSKTEPGITFLV-APFDGILGLAYPSI 189
QY 143 ARPDDSLPEPFDLSLVKQTHV-PNLFSLQCGAGFPLNOSVLAAGSGMIIGDHSYLT 201
DB 190 SASDAT--PVFDNMWNEGLVSEDLFSVYLSNG-----ERKSMVMEFGDSSYTT 237
QY 202 GSWLTPPIRREWYEVIVRVEINGQDLKM--DCKEYNDKSIIVDSTTNLRPKVFEA 259
DB 238 GSWLTPPIRREWYEVIVRVEINGQDLKM--DCKEYNDKSIIVDSTTNLRPKVFEA 291
QY 260 AVKSIKAASSTKFPDGFVLGEOLV-CWQAGTTPWNIFPVISLYLMGEVNTQSFRTIILP 318
DB 292 IGSYIGASKNL-----LGENIISCAIDSLDIVE-----TINN 325
QY 319 QQYLREVED-VATSDQDC---YKFAISQSTGT--VMGAVIMEGFYVVDVDRKRIGFAV 372
DB 326 VQYPLPASAYILKDDDDCLSGFDGMNLDTSYGLWILGDFVTRQYVTFVDFDRANNQVGLAA 385
QY 373 SA 374
DB 386 AA 387

RESULT 8
PEP4_MACFU STANDARD; PRT; 388 AA.
ID PEP4_MACFU

```

P27678;  
 01-AUG-1992 (Rel. 23, Created)  
 01-AUG-1992 (Rel. 23, Last sequence update)  
 01-NOV-1995 (Rel. 32, Last annotation update)  
 Pepsin A-4 precursor (EC 3.4.23.1) (Pepsin I/II).  
 PGA.  
 Macaca fuscata fuscata (Japanese macaque).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 Cercopithecinae; Macaca.  
 NCBI\_TaxID=9543;  
 [1]  
 SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.  
 TISSUE=Gastric mucosa;  
 MEDLINE=92037645; PubMed=1935977;  
 Kageyama T., Tanabe K., Koiwai O.;  
 "Developments-dependent expression of isozymogens of monkey  
 pepsinogens and structural differences between them.";  
 Eur. J. Biochem. 202:205-215(1991).  
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
 INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 ALSO CLEAVED TO SOME EXTENT.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPINOGEN GENES IS REGULATED BY  
 HORMONES AND RELATED SUBSTANCES.  
 CC -!- MISCELLANEOUS: EACH PEPINOGEN IS CONVERTED TO CORRESPONDING  
 PEPsin AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA  
 ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC  
 CLEAVAGE VIA AN INTERMEDIATE FORM(S).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59753; CAA42425.1; -  
 DR PIR; S16065; S16065.  
 DR PIR; S19682; S19682.  
 DR HSP; P00790; IFSN.  
 DR MEROPS; A01.001; -  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
 DR Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Multigene family; Signal; Glycoprotein.  
 FT SIGNAL 1 15  
 FT PROPEP 16 38  
 FT PROPEP 39 62  
 FT CHAIN 63 388  
 FT ACT\_SITE 94 94  
 FT ACT\_SITE 277 277  
 FT DISULFID 107 112  
 FT DISULFID 268 272  
 FT DISULFID 311 344  
 FT CARBOHYD 88 88  
 SQ SEQUENCE 388 AA; A1955 MW; A2923AB1F7FCDEB9 CRC64;  
 Query Match 14.0%; Score 302.5; DB 1; Length 388;  
 Best Local Similarity 27.6%; Pred. No. 1e-17;  
 Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;  
 QY 3 DEEPEEPRGRGSFVEMVDNLRGSGGGYVEMVGGPPPTLNTLVDTGSSNFVAGRAHP 62  
 DB 64 DEOPLE-----NVLDV-----EYFGTIGTIPAQNFVFTGSSNLWV----PSV 105

QY 63 FL-----HRYQRLSSYRDLRKGVVVPYTOCKWEGELGTDLVSPGPNVTVRANI 115  
 DB 106 YCYSLACMDHNLFPQDSSTYRATSKTVSYITGTSMTGILGYDTVRKV---GGISDTNQI 162  
 QY 116 AAITESDK-FFTINGSNWEIGILAYAEIARPDSDLSEPPFDSLVRKQTHV-PNLFSLQLCGA 173  
 DB 163 FGLSETEPGFFLYFAPDFGILGLAYPSIS--SSGATPVFDNIWQRLVSQDLFSVLSAD 220  
 QY 174 GPPLNQSEVLASVGGMIIGIDHSLYTGSLWYTPIREWYEVITIVRVEINGDL--KM 231  
 DB 221 ----DQS-----GSVIFGGIDSSYITGSLNWVPVSVEGYWQISVDSITMNGKTIACAK 270  
 QY 232 DCKEYNDKSIDVSGTTLNLRPKKVFEEAAVKSIRKAASSTKFPDGFWLGEOLV-CWQAGT 290  
 DB 271 GC-----QAIVDTGTSLLTGTPSIANIQSDICASENSD-----GENVSCSAISS 316  
 QY 291 TPWNIFPVISLYLMEVNTQSFRTILPQY-LRPVEDVATSDQCYK-----FAISQSS 344  
 DB 317 LPDIVF-----TINGVQYPLPSPAYILQSQGSCSTSGFGMDVPTEG 358  
 QY 345 TGTWGVAVIMEGFYVDFDRARKRIGFA 371  
 DB 359 ELWILGDVFIQYVTFVDFRANQVGLA 385  
 RESULT 9  
 PEPA\_CHICK  
 ID PEPA\_CHICK STANDARD; PRT; 367 AA.  
 AC P00793;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Pepsin A precursor (EC 3.4.23.1).  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 CC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=84004412; PubMed=6617663;  
 RA Baudys M., Kostka V.;  
 RT "Covalent structure of chicken pepsinogen.";  
 RL Eur. J. Biochem. 136:89-99(1983).  
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
 INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 ALSO CLEAVED TO SOME EXTENT.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 DR PIR; A00984; PECH.  
 DR HSP; P00794; 4CMS.  
 DR MEROPS; A01.001; -  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;  
 KW Glycoprotein; Gastric juice.  
 FT PROPEP 1 42  
 FT CHAIN 43 367  
 FT ACT\_SITE 77 77  
 FT ACT\_SITE 260 260  
 FT CARBOHYD 113 113  
 FT DISULFID 90 95  
 FT DISULFID 251 255  
 FT DISULFID 290 323  
 SQ SEQUENCE 367 AA; 40431 MW; 0C547E7ED8F5B341 CRC64;  
 Query Match 14.0%; Score 302; DB 1; Length 367;



	Qy	30	YYENVTVGSPPOFLNILLVDTGSSNFVCAAPHPL-----HRYTORLSSTYDLRLKG	82
	Dd	59	YYGTISITQQDQSFVIFOTGSSNLW---PSIYCKSSACSNHKREDPKSSTVYSTNET	115
	Qy	83	VVPYTOCKWEGLGTDLSIPHPNPNTVRANIAAITESDK-FFINGSNWEGILGLAYAE	141
	Dd	116	VIIAYTGSMGILGYDVAV--SSIDVQNOIGLSETEPGSFYYCNFDGIPLAEPFS	172
	Qy	142	IARPDSDLEPFDLSLVKQTHV-PNLFSLQLCGAFPLNQSEVLASVGSGMILGGIDHSLY	200
	Dd	173	IS--SSGATPVFDMMSQHVAQDLFSVYLSKDQ-----ETGSFVLFGGIDPNPT	220
	Qy	201	TGSLWYPIREWWYEVIIIVRVEINGODLK-MOCKEYNDKSVDSGTTNLRLPKKVE	258
	Dd	221	TGKIYWPVLSAETWQITMDRVTVGNKYVACFTTC-----QAIVDGTSLVMPOGAYN	274
	Qy	259	AAVKSIAKAASTE-----KFPDGFWLGEOQLVCWQAGTTPNIIPVISLYLMGEVTNOS	311
	Dd	275	RIIKDLGVSSDGEISCDDISKLPD-----VTFPHNGHA-----	307
	Qy	312	FRITILPOOYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVTRAR	365
	Dd	308	-----FTLPASAYVLNEDGSCMLGFENMGTPTELGEOWILGDVFIREYVVIEDRAN	358
	Qy	366	KRIGFA 371	
	Dd	359	NKVGLS 364	

RESULT 10  
 CATE\_HUMAN  
 D CATE\_HUMAN STANDARD; PRT; 396 AA.  
 C FI4091;  
 T 01-JAN-1990 (Rel. 13, Created)  
 T 01-JAN-1990 (Rel. 13, last sequence update)  
 T 01-MAR-2002 (Rel. 41, last annotation update)  
 E Cathepsin E precursor (BC 3.4.23.34).  
 N CTSE  
 S Homo sapiens (Human).  
 S Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 C NCBI\_TaxID=9606;  
 N [1]  
 N SEQUENCE FROM N.A.  
 N MEDLINE=89380302; PubMed=2674141;  
 P Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;  
 P "Human gastric cathepsin E. Predicted sequence, localization to  
 P chromosome 1, and sequence homology with other aspartic  
 P proteinases."; J. Biol. Chem. 264:16748-16753(1989).  
 T [2]  
 N SEQUENCE FROM N.A.  
 N MEDLINE=92112877; PubMed=1370478;  
 P Azuma T., Liu W.G., Vander Laan D.J., Bowcock A.M., Taggart R.T.;  
 P "Human gastric cathepsin E gene. Multiple transcripts result from  
 P alternative polyadenylation of the primary transcripts of a single  
 P gene locus at 1q31-q32."; J. Biol. Chem. 267:1609-1614(1992).  
 T [3]  
 N SEQUENCE FROM N.A.  
 N Tatnell P.J., Kay J.;  
 P "Human procathepsin E."; Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 T [4]  
 N SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.  
 N MEDLINE=90241267; PubMed=2334440;  
 P Athauda S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;  
 P "Structural evidence for two isozymic forms and the carbohydrate  
 P attachment site of human gastric cathepsin E."; Biochem. Biophys. Res. Commun. 168:878-885(1990).  
 T [5]

!- FUNCTION: DUE OT ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.

!- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader specificity.

!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.

!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (<http://www.isb-sib.ch/announce/>). For more information or to request a license, please contact the EBI legal department by e-mail to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).

R	EMBL; M84424; AAA52300.1; -. JOINED.
R	EMBL; M84413; AAA52300.1; JOINED.
R	EMBL; M84417; AAA52300.1; JOINED.
R	EMBL; M84418; AAA52300.1; JOINED.
R	EMBL; M84419; AAA52300.1; JOINED.
R	EMBL; M84420; AAA52300.1; JOINED.
R	EMBL; M84421; AAA52300.1; JOINED.
R	EMBL; M84422; AAA52300.1; JOINED.
R	EMBL; J05036; AAA52130.1; -. JOINED.
R	EMBL; AJ250717; CAB82850.1; -. JOINED.
R	PIR; A34401; A34401.
R	PIR; A34643; A34643.
R	PIR; A42038; A42038.
R	HSP; P00794; 4CMS.
R	MEROPS; A01.010; -. JOINED.
R	MIM; I16890; -. JOINED.
R	InterPro; IPR001969; Asp_protease.
R	InterPro; IPR001461; Pepsin.
R	Pfam; PF00026; asp; 1.
R	PRINTS; PR00792; PEPSIN.
R	PROSITE; PS00141; ASP_PROTEASE; 2.
R	HYDROLASE; Aspartyl protease; Glycoprotein; Zymogen; Signal.
KW	Signal
R	1
R	17
R	ACTIVATION PEPTIDE.
R	18
R	53
R	PROPEP
R	18
R	54
R	396
R	CHAIN
R	18
R	18
R	MOD_RES
R	96
R	96
R	ACT_SITE
R	281
R	281
R	ACT_SITE
R	60
R	60
R	DISULFID
R	109
R	114
R	DISULFID
R	272
R	276
R	DISULFID
R	314
R	351
R	CARBOHYD
R	90
R	90
R	CARBOHYD
R	220
R	220
R	CARBOHYD
R	333
R	333
R	SEQUENCE
R	396 AA; 42793 MW; 40B643C5FB01521E CRC64;

	Query Match	13.9%	Score	299.5;	DB 1;	Length	396;
	Best Local Similarity	25.9%;	Pred. No.	1.9e-178;			
	Matches	Conservative	65;	Mismatches	148;	Indels	73; Gaps
							16;
3	DEEPEECRRGSGFEMVDNLRKSGQGYVDMTVGSPPQTILNLVDTGSSNFVA	G---S	58				
	:    :	:  :      :	:				
63	DQSAREP-----LNYLD-----MEYFGTISIGSPQNFTVFDTGSSLNWVPVS	VCT	110				
	:    :	:  :      :	:				
59	AHPHPLRHRYQRGLSTYRDLRKGVYPPTYOGKEGLGTDLVSIPICPNVTVRTAAI	I	118				
	:      :	:      :	:				
111	SPACKTHSRFPQSSTYSQPQGFSIQYGTSLSGIIGADQSV--EGLTVGQGFESV	S	169				
	:    : :	:      :	:				
119	TESDKFFITGNNGWELGLAYAEIARPDSDLEPFDSLVKTHTPNLFSLQJCGAGPLN	L	178				
	:    : :	:      :	:				
170	TEPGQTFVD--AEFDGILLGLGPSLA--VGGVTPTVFDNNMAQ----	NLYDLPMSFYIMSN	222				
	:    : :	:      :	:				
179	OSEVLASVGGSMITGDHSLYTSGSLWTPTFRREWWYVIIVRVIEINGODLKMDCKE	VNY	238				
	:      :	:      :	:				
223	PR--GGAGSELIFGYDHSHFSGSLNWVPVKTKAQYQLNDNIQVGG--TWMFCE-	-G	275				





Pfam: PF00026; asp: 1;  
 DR PRINTS; PRO0792; PEPSIN.  
 DR PROSITE; PS00141; ASP-PROTEASE; 2.  
 DR Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Signal; Phosphorylation; Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 59 ACTIVATION PEPTIDE.  
 FT CHAIN 60 387 PEPSIN II-2/3.  
 FT MOD\_RES 129 129 PHOSPHORYLATION (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 276 276 BY SIMILARITY.  
 FT ACT\_SITE 106 111 BY SIMILARITY.  
 FT DISULFID 267 271 BY SIMILARITY.  
 FT DISULFID 310 343 BY SIMILARITY.  
 FT SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;  
 Query Match 13.7%; Score 296; DB 1; Length 387;  
 Best Local Similarity 26.9%; Pred. No. 3.5e-17; Indels 66; Gaps 13;  
 Matches 97; Conservative 63; Mismatches 135;  
 QY 30 YVEMTGVSPQTLNVLVDGTSNFAVGAAPHPF-----LHRYQRLSSYRDLRKG 82  
 DB 75 YFTISIGTPQDFVIFDTGSSNLV---PSTYCSSLACALHKRNPEDSSTYQGTSET 131  
 QY 83 VYVPYTOGKWEGLGTLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEI 142  
 DB 132 LSITYTGSMTGILGYDVKVGSIEDTQIFGLSKTEPSLTLF--APPDGLGLATPSI 189  
 QY 143 ARPDSLEPFDLSLVKQTHV--PNFLSLOLCGAGFPLNQSSEVLASVGGSMIIGGDHSLYT 201  
 DB 190 SSSDAT--PVFNMMNEGILVSQDLFSVYLSSD-----EKGSLVMFGGIDSSVYT 237  
 QY 202 GSLWTPPIREYEVIIIVRVEINGODLKM--DKEYNVDKSYDSGTNLRPKKVFPEA 259  
 DB 238 GSLWVVPVSEGYWQITMDSVINGETIACADSC-----QAIVDTGSLTGP-----TS 287  
 QY 260 AVKSIKAASSTKFPDGFNLGQV--CMQAGTTPNIPFPVSLYLMGEVTVNQSFRTILP 318  
 DB 288 ATSNISQYIGASK-----NLGGENVISCASIDSLDIVF-----TING 325  
 QY 319 QQVLRVEDVATSDQDCYKFAISQSSTGT-----VMGAVIMEGYVYVDFDRARRIGAV 372  
 DB 326 IQYPLPASAYLKEDDCTSGLEGMMVDVYTGELMTLGDVFIQYFTVDFRANNOGLGAA 385  
 QY 373 S 373  
 DB 386 A 386  
 RESULT 13  
 CATE\_CAVPO STANDARD; PRT; 391 AA.  
 ID CATE\_CAVPO  
 AC P25796;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cathepsin E precursor (EC 3.4.23.34).  
 GN CrSE.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystricognathia; Caviidae; Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=92355614; PubMed=1644829;  
 RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koizumi O.,  
 RA Tanji M., Yakabe E., Achauda S.B., Takahashi K.;  
 RT "Gastric procathepsin E and procathepsin from guinea pig.  
 RT Purification, molecular cloning of cDNAs, and characterization of  
 RT enzymatic properties, with special reference to procathepsin E";  
 RL J. Biol. Chem. 267:16450-16459(1992).  
 RN [2]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Gastric mucosa;  
 RX MEDLINE=96073637; PubMed=8540321;  
 RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,  
 RA Tanji M., Achauda S.B., Takahashi K.;  
 RT "Isolation, characterization, and structure of procathepsin E and  
 RT cathepsin E from the gastric mucosa of guinea pig."  
 FT SIGNAL 1 15  
 FT PROPEP 16 59 ACTIVATION PEPTIDE.  
 FT CHAIN 60 387 PEPSIN II-2/3.  
 FT MOD\_RES 129 129 PHOSPHORYLATION (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 276 276 BY SIMILARITY.  
 FT ACT\_SITE 106 111 BY SIMILARITY.  
 FT DISULFID 267 271 BY SIMILARITY.  
 FT DISULFID 310 343 BY SIMILARITY.  
 FT SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;  
 Query Match 13.7%; Score 296; DB 1; Length 391;  
 Best Local Similarity 26.9%; Pred. No. 3.6e-17; Indels 72; Gaps 16;  
 Matches 98; Conservative 64; Mismatches 130;  
 QY 30 YVEMTGVSPQTLNVLVDGTSNFAVGA---APHPFLHRYQRLSSYRDLRKGYYV 85  
 DB 74 YFTISIGSPQNTFTVIFDTGSSNLVPSVYCTSPACQTHPVFHPSLSSYREYVNSFSI 133  
 QY 86 PYTOGKWEGLGTLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIAR 145  
 DB 134 QXGTGSLTGIIGADQSV--EGLTVGQGFGEVQEPGKTFVH--AEFDGLGLGTPSLAA- 190  
 QY 146 DDLSEPFDSLVKQTHVNPFLSLOLCGAGFPLNQSSEVLASVGGSMIIGGDHSLYTSLW 205  
 DB 191 -GGVTPVFDNMAQ---NLVALPM---FSVYMSNPGSGSELTFGYPDFSHFSGSLN 241  
 QY 206 YTPIRREYEVIIIVRVEINGODLKMCKEYNDKSYDSGTNLRPKKVFPEAAVRSIK 265  
 DB 242 WVPVTKAQWQIALDGIQV---DSVMFCSE--GCQAVDTGTSLTGP---PGKIKQLQ 293  
 QY 266 AASSTKFPDGFNLGQVCHQAGTTPNIPFPVSLYLMGEVTVNQSFRI---TILPQQ 320  
 DB 294 EALGATYVDEGY-----SVQC-----ANLNMMLDVT---FIINGVPTLNPTA 333  
 QY 321 YLRPVEDVATSDQDCYKFAISQSSTG-----TVMGAVIMEGYVYVDFDRARR 367

```

Db      334 Y--TLDFVDMQVC-----STGFEGLIOPAGPLWILGDVIRFYAVFDGRNMR 383
QY      368 IGFA 371
Db      384 VGLA 387

RESULT 14
CATD_HUMAN
ID      CATD_HUMAN      STANDARD;      PRT;      412 AA.
AC      P07339;
DT      01-APR-1988 (Rel. 07, Created)
DT      01-APR-1988 (Rel. 07, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Cathepsin D precursor (EC 3.4.23.5).
GN      CTSD.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX      MEDLINE=85270436; PubMed=3927292;
RA      Faust P.L., Kornfeld S., Chirgwin J.M.;
RT      "Cloning and sequence analysis of cDNA for human cathepsin D.";
RL      Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).
[2]
SEQUENCE FROM N.A.
RX      MEDLINE=87231068; PubMed=3588130;
RA      Westley B.R., May F.E.B.;
RT      "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive
RT      human breast cancer cells.";
RL      Nucleic Acids Res. 15:3773-3786(1987).
[3]
SEQUENCE FROM N.A.
RX      MEDLINE=91299158; PubMed=2069717;
RA      Redeker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;
RT      "Molecular organization of the human cathepsin D gene.";
RL      DNA Cell Biol. 10:423-431(1991).
[4]
SEQUENCE OF 1-22 FROM N.A.
RX      MEDLINE=94085791; PubMed=8262386;
RA      May F.E., Smith D.J., Westley B.R.;
RT      "The human cathepsin D-encoding gene is transcribed from an estrogen-
RT      regulated and a constitutive start point.";
RL      Gene 134:277-282(1993).
[5]
SEQUENCE OF 1-22 FROM N.A.
RX      MEDLINE=95021301; PubMed=7935485;
RA      Augereau P., Miralles F., Cavaillès V., Gaudeliet C., Parker M.,
RA      Rochefort H.;
RT      "Characterization of the proximal estrogen-responsive element of
RT      human cathepsin D gene.";
RL      Mol. Endocrinol. 8:693-703(1994).
[6]
SEQUENCE OF 170-180.
RX      TISSUE=Liver;
RA      Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
RA      Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
RA      Appel R.D., Hughes G.J.;
RL      Submitted (JUN-1992) to the SWISS-PROT data bank.
[7]
VARIANT VAL-58.
RX      MEDLINE=20179010; PubMed=10716266;
RA      Papassotiropoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,
RA      Maier W., Pauls J., Lautenschlager N., Heun R.;
RT      "A genetic variation of cathepsin D is a major risk factor for
RT      Alzheimer's disease.";
RL      Ann. Neurol. 47:399-403(2000).
[8]
X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RC      TISSUE=Spleen;

```

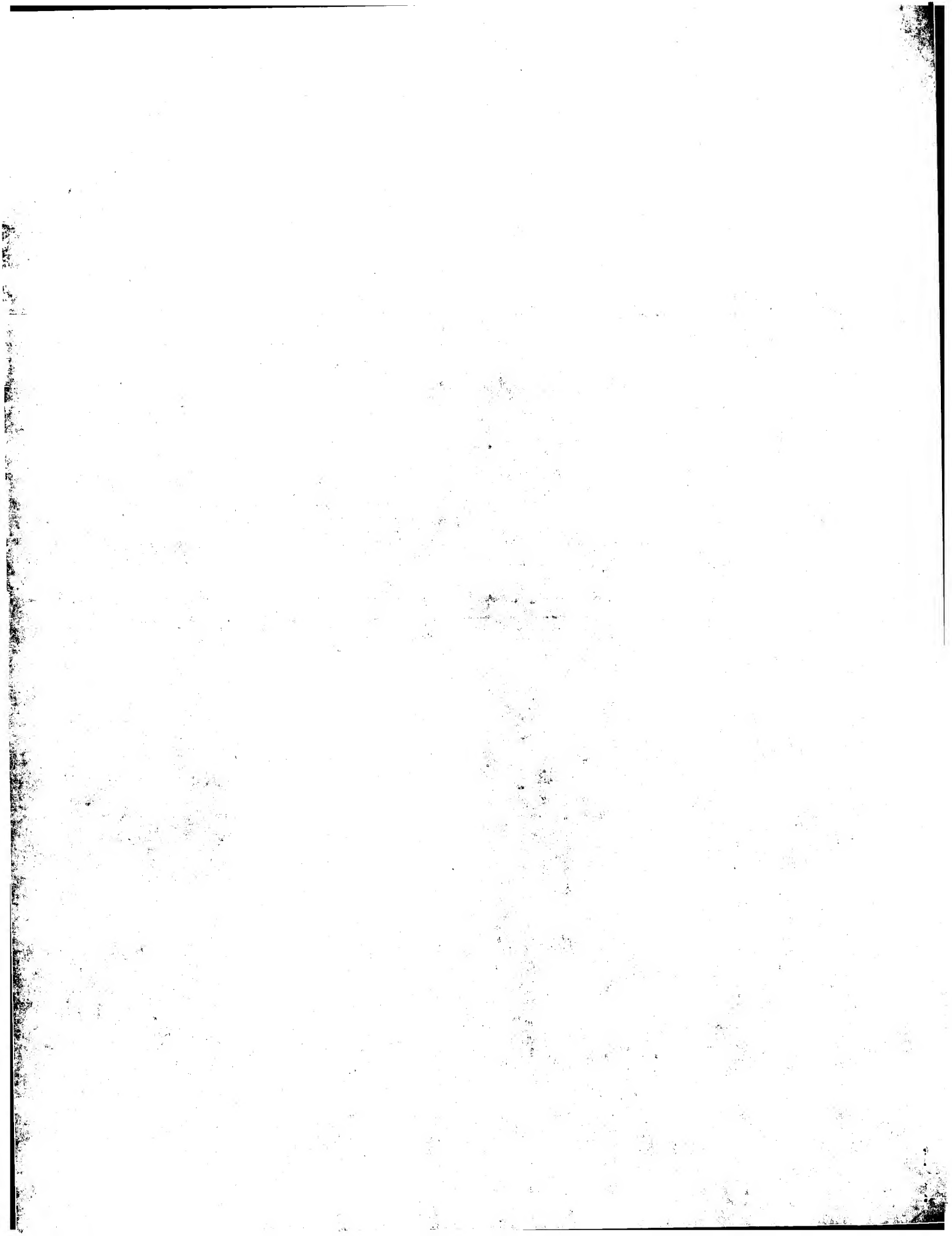
```

RX      MEDLINE=93223670; PubMed=8467789;
RA      Metcalf P., Fusek M.;
RT      "Two crystal structures for cathepsin D: the lysosomal targeting
RT      signal and active site.";
RL      EMBO J. 12:1293-1302(1993).
[9]
RN      X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RC      TISSUE=Liver;
RX      MEDLINE=93342076; PubMed=8393577;
RA      Baldwin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,
RA      Cachau R.E., Collins J., Silva A.M., Erickson J.W.;
RT      "Crystal structures of native and inhibited forms of human cathepsin
RT      D: Implications for lysosomal targeting and drug design.";
RL      Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).
CC      -!- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
CC      PROTEIN BREAKDOWN.
CC      -!- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
CC      that of pepsin A. Does not cleave the 4-Gln-1-His-5 bond in B
CC      chain of insulin.
CC      -!- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC      -!- SUBCELLULAR LOCATION: Lysosomal.
CC      -!- POLYMORPHISM: The Val-58 allele is significantly overrepresented
CC      in demented patients (11.8%) compared with nondemented controls
CC      (4.9%). Carriers of the Val-58 allele have a 3.1-fold increased
CC      risk for developing AD than noncarriers.
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC      EUKARYOTIC ASPARTYL PROTEASES FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M11233; AAB59529.1; -
EMBL; X05344; CAA28955.1; -
EMBL; M63138; AAA51922.1; -
EMBL; M63134; AAA51922.1; JOINED.
EMBL; M63135; AAA51922.1; JOINED.
EMBL; M63136; AAA51922.1; JOINED.
EMBL; M63137; AAA51922.1; JOINED.
EMBL; L12980; AAA16314.1; -
EMBL; S74689; AAD14156.1; -
EMBL; S52557; AAD13868.1; -
PIR; A25771; KHHUD.
PDB; 1LYA; 31-JAN-94.
PDB; 1LYB; 31-JAN-94.
MEROPS; A01.009; -
SWISS-2DPAGE; P07339; HUMAN.
Siena-2DPAGE; P07339; -
MIM; 116840; -
InterPro; IPR001969; Asp_protease.
InterPro; IPR001461; Pepsin.
Pfam; PF00026; asp; 1.
PRINTS; PR00792; PEPsin.
PROSITE; PS00141; ASP_PROTEASE; 2.
KW      Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Signal; Zymogen;
KW      Polymorphism; Alzheimer's disease; 3D-structure.
FT      SIGNAL 1 18
FT      PROPEP 19 64      ACTIVATION PEPTIDE.
FT      CHAIN 65 412      CATHEPSIN D.
FT      CHAIN 169 412      LIGHT CHAIN (PROBABLE).
FT      CHAIN 169 412      HEAVY CHAIN (PROBABLE).
FT      ACT_SITE 97 97
FT      ACT_SITE 295 295
FT      DISULFID 91 160
FT      DISULFID 110 117
FT      DISULFID 286 290
FT      DISULFID 329 366
FT      CARBOHYD 134 134
FT      CARBOHYD 263 263
FT      N-LINKED (GLCNAC. . .).
FT      N-LINKED (GLCNAC. . .).

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:17:12 ; Search time 50.29 Seconds  
(without alignments)  
1400.058 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEPEEPGRGSGFVEMVD.....GPFVTLDMEDCGYNIPOTDE 407

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2156	100.0	532	4 Q9ULS1	Q9uls1 homo sapien
2	2001.5	92.8	476	4 Q9BYC1	Q9byc1 homo sapien
3	1898	88.0	457	4 Q9BYC0	Q9byc0 homo sapien
4	1753.5	81.3	432	4 Q9BYB9	Q9byb9 homo sapien
5	1149	53.3	266	11 Q9CUU5	Q9cuu5 mus musculus
6	1121	52.0	439	4 Q9HZV8	Q9hzv8 homo sapien
7	1116.5	51.8	514	11 Q9JL18	Q9jl18 mus musculus
8	967.5	44.9	396	4 Q9NZL1	Q9nzl1 homo sapien
9	933	43.3	468	4 Q9NZL2	Q9nzl2 homo sapien
10	712.5	33.0	213	4 Q9F0D2	Q9f0d2 homo sapien
11	566.5	26.3	255	11 Q9R1P7	Q9rlp7 mus musculus
12	332.5	15.4	391	5 Q9VKP6	Q9vkep6 drosophila
13	332	15.4	354	5 Q9GYX7	Q9gyx7 boophilus m
14	312.5	14.5	386	6 Q9BGU5	Q9bgus bos taurus
15	308	14.3	384	13 Q9DEC2	Q9dec2 xenopus lae
16	308	14.3	385	13 Q9DEC4	Q9dec4 rana catesb

17	305	14.1	386	6 Q9GMV7	Q9gmv7 rhinolophus
18	305	14.1	387	6 Q9GMV8	Q9gmv8 sorex ungui
19	304.5	14.1	372	5 Q9VLK3	Q9vilk3 drosophila
20	304.5	14.1	387	13 Q9DDV5	Q9ddv5 salvelinus
21	304	14.1	387	6 Q9GMV9	Q9gmv9 suncus muri
22	303.5	14.1	383	13 Q9DE45	Q9de45 salvelinus
23	303	14.1	383	13 Q9DEC3	Q9dec3 xenopus lae
24	302.5	14.0	376	13 Q9PUR8	Q9pur8 pseudopleur
25	301	14.0	382	13 Q9PRG9	Q9prg9 gallus gall
26	301	14.0	423	5 Q9VKP7	Q9vkep7 drosophila
27	298.5	13.8	384	13 Q9L322	Q9l322 rana catesb
28	295.5	13.7	386	6 Q9GMV6	Q9gmv6 canis fami
29	294	13.6	396	13 Q93428	Q93428 chionodraco
30	290.5	13.5	381	6 Q9GK11	Q9gk11 camelus dro
31	288	13.4	399	13 Q93458	Q93458 podarcis si
32	287.5	13.3	444	5 Q21966	Q21966 caenorhabdl
33	284	13.2	398	13 P87370	P87370 oncorhynch
34	284	13.2	427	5 P91802	P91802 schistosoma
35	281	13.0	378	13 Q9P0R9	Q9p0r9 pseudopleur
36	279.5	13.0	390	6 Q9GK10	Q9gk10 camelus dro
37	278	12.9	370	6 Q9TTW1	Q9ttw1 bos taurus
38	278	12.9	399	13 Q9DD89	Q9dd89 brachydantio
39	278	12.9	422	5 Q96906	Q96906 onchocerca
40	277.5	12.9	380	6 Q28950	Q28950 sus scrofa
41	277	12.8	446	5 Q9N9H3	Q9n9h3 necator ame
42	276	12.8	389	13 Q9PWK1	Q9pwk1 gallus gall
43	276	12.8	389	13 Q9W643	Q9w643 gallus gall
44	276	12.8	396	13 Q9DEX3	Q9dex3 clupea hare
45	275	12.8	385	6 Q29080	Q29080 sus scrofa

#### ALIGNMENTS

RESULT 1  
Q9ULS1 PRELIMINARY; PRT; 532 AA.  
ID Q9ULS1  
AC Q9ULS1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1149 PROTEIN (FRAGMENT).  
GN KIAA1149.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20039618; PubMed=10574461;  
RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;  
RT "Characterization of cDNA clones selected by the GeneMark analysis  
RT from size-fractionated cDNA libraries from human brain.";  
RL DNA Res. 6:329-336(1999).  
DR EMBL; AB032975; BAA86463.2; .  
DR HSSP; P56272; 1AM5.  
DR MEROPS; A01.004; .  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 532 AA; 58720 MW; 98B135D0D5FBD2E8 CRC64;

Query Match 100.0%; Score 2156; DB 4; Length 532;  
Best Local Similarity 100.0%; Pred. No. 3.3e-175;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLRGKSGGYVEMTVGSPQDTLNLIVDTGGSNFAVGAAP 60  
|||||

```

Db 77 ETDEEPEEPCRRGSFVEMVDNLKSGQGYVEMTVGSPQTLNLLVDTGSSNFVAGAAP 136
QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQCKWEGELGTDLSIPHGPNTVVRANIAAITE 120
Db 137 HPFLHRYQRLSSTYRDLRKGVYVPTQCKWEGELGTDLSIPHGPNTVVRANIAAITE 196
QY 121 SDRFFINGSNWEGILGLAYAEIARPDLSLVKQTHVFNLSQLCGAGFPLNQS 180
Db 197 SDRFFINGSNWEGILGLAYAEIARPDLSLVKQTHVFNLSQLCGAGFPLNQS 256
QY 181 EVLASVGSMIIGIDHSLYTGSMTPIRREWYEVIIIVRVEINGQDLKMDCKEYNIDK 240
Db 257 EVLASVGSMIIGIDHSLYTGSMTPIRREWYEVIIIVRVEINGQDLKMDCKEYNIDK 316
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVLGEQVLCVQAGTTPWNIFPVIS 300
Db 317 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVLGEQVLCVQAGTTPWNIFPVIS 376
QY 301 LYLMEVNTQSFRTITLPOQYLRVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 377 LYLMEVNTQSFRTITLPOQYLRVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 436
QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 407
Db 437 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 483
RESULT 2
Q9BYC1 PRELIMINARY; PRT; 476 AA.
ID Q9BYC1
AC Q9BYC1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-476.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production."
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050436; BAB40931.1; -.
DR HSSP; P32329; LYPS.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 476 AA; 52907 MW; 6C8C87F8A953AF66 CRC64;

Query Match 92.8%; Score 2001.5; DB 4; Length 476;
Best Local Similarity 93.9%; Pred. No. 4.le-162;
Matches 382; Conservative 0; Mismatches 0; Indels 25; Gaps 1;

QY 1 ETDEEPEEPCRRGSFVEMVDNLKSGQGYVEMTVGSPQTLNLLVDTGSSNFVAGAAP 60
Db 46 ETDEEPEEPCRRGSFVEMVDNLKSGQGYVEMTVGSPQTLNLLVDTGSSNFVAGAAP 105
QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQCKWEGELGTDLSIPHGPNTVVRANIAAITE 120
Db 106 HPFLHRYQRLSSTYRDLRKGVYVPTQCKWEGELGTDLSIPHGPNTVVRANIAAITE 165
QY 121 SDRFFINGSNWEGILGLAYAEIARPDLSLVKQTHVFNLSQLCGAGFPLNQS 180

```

```

Db 166 SDKFFINGSNWEGILGLAYAEIARPDLSLVKQTHVFNLSQLCGAGFPLNQS 200
QY 181 EVLASVGSMIIGIDHSLYTGSMTPIRREWYEVIIIVRVEINGQDLKMDCKEYNIDK 240
Db 201 EVLASVGSMIIGIDHSLYTGSMTPIRREWYEVIIIVRVEINGQDLKMDCKEYNIDK 260
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVLGEQVLCVQAGTTPWNIFPVIS 300
Db 261 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFVLGEQVLCVQAGTTPWNIFPVIS 320
QY 301 LYLMEVNTQSFRTITLPOQYLRVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 321 LYLMEVNTQSFRTITLPOQYLRVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 380
QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 407
Db 381 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPQTD 427
RESULT 3
Q9BYC0 PRELIMINARY; PRT; 457 AA.
ID Q9BYC0
AC Q9BYC0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME
DE TYPE C).
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production."
RL Neurosci. Lett. 307:9-12(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EXOCRINE PANCREAS;
RA Zaccchetti D., De Pietri Tonelli D., Schurbus R.;
RT "New beta-site APP cleaving enzyme isoform (BACE-1C) obtained from
RT human pancreas."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050437; BAB40932.1; -.
DR EMBL; AF338817; AAK38375.1; -.
DR HSSP; P32329; LYPS.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 457 AA; 51068 MW; C794C9A9E85FE7A2 CRC64;

Query Match 88.0%; Score 1898; DB 4; Length 457;
Best Local Similarity 89.2%; Pred. No. 2.5e-153;
Matches 363; Conservative 0; Mismatches 0; Indels 44; Gaps 1;

QY 1 ETDEEPEEPCRRGSFVEMVDNLKSGQGYVEMTVGSPQTLNLLVDTGSSNFVAGAAP 60
Db 46 ETDEEPEEPCRRGSFVEMVDNLKSGQGYVEMTVGSPQTLNLLVDTGSSNFVAGAAP 105
QY 61 HPFLHRYQRLSSTYRDLRKGVYVPTQCKWEGELGTDLSIPHGPNTVVRANIAAITE 120
Db 106 HPFLHRYQRLSSTYRDLRKGVYVPTQCKWEGELGTDLSIPHGPNTVVRANIAAITE 145
QY 121 SDRFFINGSNWEGILGLAYAEIARPDLSLVKQTHVFNLSQLCGAGFPLNQS 180

```



```

Db 146 -----PDSLEFFSLVKQTHVPLNLSLQCGAGFLNQS 181
QY 181 EVLASVGSMIIGDHSYTGSLWYTPIRREWYIEVIRVEINGQDLKMDCKEYNNDK 240
Db 182 EVLASVGSMIIGDHSYTGSLWYTPIRREWYIEVIRVEINGQDLKMDCKEYNNDK 241
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVIS 300
Db 242 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVIS 301
QY 301 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 302 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 361
QY 361 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPOTDE 407
Db 362 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPOTDE 408

RESULT 4
Q9BYB9
ID Q9BYB9 PRELIMINARY; PRT; 432 AA.
AC Q9BYB9;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-432.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production.";
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050438; BAB40933.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp_protease.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 432 AA; 48212 MW; 96FC81EGF0EED01B CRC64;

Query Match 81.3%; Score 1753.5; DB 4; Length 432;
Best Local Similarity 83.0%; Pred. No. 4.8e-141;
Matches 338; Conservative 0; Mismatches 0; Indels 69; Gaps 1;

QY 1 ETDEPEPPGRGFSFVEMVNDLRKSGQYVEMTVGSPPTNLNLDVTGSSNFVAGAAP 60
Db 46 ETDEPEPPGRGFSFVEMVNDLRKSGQYVEMTVGSPPTNLNLDVTGSSNFVAGAAP 105
QY 61 HPFLHRYQRLSTYRDLRGVYVPTQKWEGLGTLDLVIPHGNVTVRANIAAITE 120
Db 106 HPFLHRYQRLSTYRDLRGVYVPTQKWEGLGTLDLVIPHGNVTVRANIAAITE 120
QY 121 SDKFFINGSNWEGILGAYAEIARPDSDLEPFDLSLVKQTHVPLNLSLQCGAGFLNQS 180
Db 146 -----LCGAGFLNQS 156
QY 181 EVLASVGSMIIGDHSYTGSLWYTPIRREWYIEVIRVEINGQDLKMDCKEYNNDK 240
Db 157 EVLASVGSMIIGDHSYTGSLWYTPIRREWYIEVIRVEINGQDLKMDCKEYNNDK 216
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVIS 300

```

```

Db 217 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVIS 276
QY 301 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 277 LYLMEVNTNQSFRITILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYV 336
QY 361 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPOTDE 407
Db 337 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPOTDE 383

RESULT 5
Q9CUU5
ID Q9CUU5 PRELIMINARY; PRT; 266 AA.
AC Q9CUU5;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME (FRAGMENT).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=BRAIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Cariboldi M.,
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willing L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014390; BAB29317.1; -.
DR MEROPS; A01.004; -.
DR MGD; MGI:1346542; Bace.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 2.
FT NON_TER 1
SQ SEQUENCE 266 AA; 30201 MW; B913FDA8ADAB4238 CRC64;

Query Match 53.3%; Score 1149; DB 11; Length 266;
Best Local Similarity 99.1%; Pred. No. 7.7e-90;
Matches 215; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 191 IIGGIDHSYTGSLWYTPIRREWYIEVIRVEINGQDLKMDCKEYNNDKSIDVSGTTNL 250
Db 1 IIGGIDHSYTGSLWYTPIRREWYIEVIRVEINGQDLKMDCKEYNNDKSIDVSGTTNL 60
QY 251 RLPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVISLMEGVNTQ 310
Db 61 RLPKKVFEEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVISLMEGVNTQ 120
QY 311 SFRITILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVDFRARRKRG 370
Db 121 SFRITILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVDFRARRKRG 180

```



```
RP SEQUENCE FROM N.A.
RX MEDLINE-20422477; PubMed-10965118;
RA Solans A., Estivill X., de La Luna S.;
RT "A new aspartyl protease on 21q22.3; BACE2, is highly similar to
RL Alzheimer's amyloid precursor protein beta-secretase.";
DR EMBL; AF188277; AAF35836.1; -.
DR HSSP; P00797; 2REN.
DR MEROPS; A01.041; -.
DR InterPro; IP0001969; Asp_protease.
DR Pfam; PF00026; asp; 2.
DR PRINTS; PR00792; Pepsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

  Query Match      44.9%; Score 967.5; DB 4; Length 396;
  Best Local Similarity 54.6%; Pred. No. 4e-74;
  Matches 183; Conservative 54; Mismatches 85; Indels 13; Gaps 4;

QY 9 PGRGSEFVEMVNDLRGKSGQGYVYVMTGSPPTNLILVDTGSSNFVAAGAPFLHRY 68
Db 71 PAGAANFLAMVDNLQDGRGYYLEMLIGTPQKQLILVDTGSSNFVAGTPHSDITYF 130
QY 69 QRLSSTYRDLRKGVYVYPTQGWKEGELGTLVSIPIHGPNTVTRANIAAATESDKFFNG 128
Db 131 DTERSSTYRSKGFDTVKYTOGSMFTGVEDLVTIPKGFNTSFLVNIATIFESNFPLPG 190
QY 129 SNWEGILGLAYAEIARPDLSLEPFDSLVKQTHVNPFLSLOLCGAGFLNQSSEVLASVG 188
Db 191 IKWNGILGLAYATLAKPSSLETFFDSLVQTQANIPNVFSMQMCGAGLPVAGS---GTNGG 247
QY 189 SMIIIGDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNDKSIDVSGTT 248
Db 248 SLVGGIEPSLYKGDWYTPIKEWYQIETLKEIGGSLNLCREYNADKAIVDSGTT 307
QY 249 NLRPKVFEAAVSKIAASSTKTEKPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMEV 308
Db 308 LLRLPKQVDAVAVARASLPEFSDGFWTGSOLACWTNSETPSWYFPKISIVLRDENS 367
QY 309 NOSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVYFDRARKRI 368
Db 368 SRSFRITILPQK-LRVLQ-----CLKPFLQSQ 393

RESULT 9
Q9NZL2 ID Q9NZL2 PRELIMINARY; PRT; 468 AA.
AC Q9NZL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20422477; PubMed-10965118;
RA Solans A., Estivill X., de La Luna S.;
RT "A new aspartyl protease on 21q22.3; BACE2, is highly similar to
RL Alzheimer's amyloid precursor protein beta-secretase.";
DR EMBL; AF188276; AAF35835.1; -.
DR HSSP; P00797; 2REN.
DR MEROPS; A01.041; -.
DR InterPro; IP0001969; Asp_protease.
DR InterPro; IP0001461; Pepsin.
KW Pfam; PF00026; asp; 3.
```

```
DR PRINTS; PR00792; Pepsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

  Query Match      43.3%; Score 933; DB 4; Length 468;
  Best Local Similarity 46.5%; Pred. No. 4.5e-71;
  Matches 180; Conservative 59; Mismatches 94; Indels 54; Gaps 3;

QY 9 PGRGSEFVEMVNDLRGKSGQGYVYVMTGSPPTNLILVDTGSSNFVAAGAPFLHRY 68
Db 71 PAGAANFLAMVDNLQDGRGYYLEMLIGTPQKQLILVDTGSSNFVAGTPHSDITYF 130
QY 69 QRLSSTYRDLRKGVYVYPTQGWKEGELGTLVSIPIHGPNTVTRANIAAATESDKPEING 128
Db 131 DTERSSTYRSKGFDTVKYTOGSMFTGVEDLVTIPKGFNTSFLVNIATIFESNFPLPG 190
QY 129 SNWEGILGLAYAEIARPDLSLEPFDSLVKQTHVNPFLSLOLCGAGFLNQSSEVLASVG 188
Db 191 IKWNGILGLAYATLAKPSSLETFFDSLVQTQANIPNVFSMQMCGAGLPVAGS---GTNGG 247
QY 189 SMIIIGDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNDKSIDVSGTT 248
Db 248 SLVGGIEPSLYKGDWYTPIKEWYQIETLKEIGGSLNLCREYNADKAIVDSGTT 307
QY 249 NLRPKVFEAAVSKIAASSTKTEKPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMEV 308
Db 308 LLRLPKQVDAVAVARASLPEFSDGFWLGEQLVCWQAGTTPWNIFPVISLYLMEV 376
QY 309 NOSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVYFDRARKRI 368
Db 330 -----YIQPMGAGLNY-ECYRFGISPTNALVIGATVMGEGFYVIFDRAQKRV 376
QY 369 GFAVSACHVHDEFRTAAVEGPFVTLDM 395
Db 377 GFAASPCAEIAGAAVSEISGPFSTEDV 403

RESULT 10
Q9P0D2 ID Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HSPCL04 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1; -.
DR InterPro; IP0001461; Pepsin.
DR Pfam; PF00026; asp; 1.
FT NON_TER 1
SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;

  Query Match      33.0%; Score 712.5; DB 4; Length 213;
  Best Local Similarity 83.5%; Pred. No. 8.7e-53;
  Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

QY 193 GGIDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNDKSIDVSGTTNRL 252
Db 1 GGIDHSLSYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNDKSIDVSGTTNRL 60
```

QY 253 PKKVEAAVKSIIKAASSTKFPDGLGQLVCMQAGTTPWNIFFVISLYLMGEVTVNSQF 312  
Db 61 PKKVEAAVKSIIKAASSTKFPDGLGQLVCMQAGTTPWNIFFVISLYLMGEVTVNSQF 120  
QY 313 RIITLPOQYLRPVEDVATSDODCYKFAISQSGTGTVMGAVIMEG 356  
Db 121 RIITLPOQYLRP-----WKMPRPKTTVTVCCHLTIVHG 153  
RESULT 11  
Q9RLP7 PRELIMINARY; PRT; 255 AA.  
AC Q9RLP7; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE ASPARTYL PROTEASE (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A.  
RA Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;  
RA "Cloning of a gene from chromosome 21 Down Region encoding a potential  
RA transmembrane protease.";  
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AF051150; AAD45964.1; -  
DR MEROPS; A01.041; -  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF000026; asp; 2.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
KW protease.  
FT NON\_TER  
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;  
Query Match 26.3%; Score 566.5; DB 11; Length 255;  
Best Local Similarity 53.4%; Pred. No. 3.1e-40;  
Matches 102; Conservative 35; Mismatches 53; Indels 1; Gaps 1;  
QY 205 WYTPTRREWYVIVRVVEINGQDLKMDCKEYNDKSIDVSGTTLNLRPKKVFEEAAVSKI 264  
Db 1 WYTPKEEYVYIETLKLKGQNLNLDREYNADKAIVDSGTTLLRLPQKVFDAVVEAV 60  
QY 265 KAASSTKFPDGLGQLVCMQAGTTPWNIFFVISLYLMGEVTVNSQFRIITLPOQYLRP 324  
Db 61 ARTSLIPEFSDGFWTGAQLACWTNSETPWAYPKISIVLRDENDASRPRITILPOLYIOP 120  
QY 325 VEDVATSDODCYKFAISQSGTGTVMGAVIMEGFYVVFDRARKRIGFAYSVACHVHDFRPTA 384  
Db 121 MMGAGNY-ECYRFGIISSTNALVIGATVMGSEFYVVFDRQRRVGFVSPCAEIEGTTVS 179  
QY 385 AVEGPEVTLDL 395  
Db 180 EISGPESTEDI 190  
RESULT 12  
Q9VKP6 PRELIMINARY; PRT; 391 AA.  
AC Q9VKP6; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE CG17134 PROTEIN.  
GN CG17134.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnikier S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,  
RA George R.A., Lewis S.E., Zhang Q., Chen L.X.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Pfeiffer B.D.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej D.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Rubin G.M., Venter J.C.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR EMBL; AE003630; AAF53016.1; -  
DR HSP; P00794; 4CMS.  
DR MEROPS; A01.0PW; -  
DR FlyBase; FBgn0032304; CG17134.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF000026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
DR Aspartyl protease; Hydrolase.  
KW SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;  
Query Match 15.4%; Score 332.5; DB 5; Length 391;  
Best Local Similarity 29.1%; Pred. No. 5.1e-20;  
Matches 105; Conservative 61; Mismatches 146; Indels 49; Gaps 13;  
QY 20 DNLRGKGGYVEMTVGSPPTLNLIVDTGSSNFVGAAPHPF-----LHRYQRQLS 73  
Db 66 ENLHNSMNEYGVIAIGTPEQRNFILFTGSAANLWPSASCPASNTACQRHNYDSAS 125  
QY 74 STYDLRGVYVPTQCKWEGELGTLVSIPIGPNVTYRANLAATESDKFFINGSNWEG 133  
Db 126 STYVANGEEFAEYGTGSLSGFLSNDIVTIA-GISIQNTGELSEPTTFVD-APFAG 183  
QY 134 ILGLAYAEIARPDSDLEPFFDSLVKQTHVPN-LFSLQLCGAGFPLNQSEVLASVGGSMII 192  
Db 184 ILGLAFSAIA--VDGVTTPFDNMISQGLLDEPVISFYLKROG-----TAVRGELIL 233

```
QY 193 GGIDHSLYTGSLWYTPIRREWYEVILVRVEINGODLMDCKEYNDKSIYDVGTTNLR 252
||||| ||| ||| ||| ||| : : : : : ||| : : : ||| : : : ||| : : :
Db 234 GGIDSLYRGSLTYVPVPAVKVNTIKNTGILLNGC-----QAIDTGTSLIAY 287
QY 253 PKKVEAAVKSKAASSTKEKPDGFWLGEOLY-CWAGTTPWNIPFPVLSLYLMGEVNTQS 311
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 288 PLAAAYRKINRQLGATDND-----GEAFVRCGRVSS-----LPKYNLNLIGTV---- 329
QY 312 FRITILPOQYLRPEVDVATSDQDCYK-FAISQSSTGTVGAVIMEGFVYVDRARKRIGF 370
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 330 --FTLAPRDYI--VKVTONGOTYCMSAFTYMEGLSFWLGDVFIKGYTFVDFDKGNERIGF 385
QY 371 A 371
Db 386 A 386

RESULT 13
Q9GYX7 PRELIMINARY; PRT; 354 AA.
AC Q9GYX7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE HEME-BINDING ASPARTIC PROTEINASE (FRAGMENT).
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Boophilidae; Boophilus.
OX NCBI_TaxID=6941;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PORTO ALEGRE; TISSUE=OVARY;
RA Sorgine M.H.F., Logullo C., Zingali R.B., Paiva-Silva G.O.,
RA Juliano L., Olivella P.L.;
RT "A hemo-binding aspartic proteinase from the eggs of the hard tick
RT Boophilus microplus."
RL J. Biol. Chem. 0:0-0(2000).
DR EMBL; AF286865; AAG00993.1;
DR HSP; P00797; 2REN.
DR MEROPS; A01.054;
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp.1.
DR PRINTS; PR00792; PEPsin.
FT NON_TER 1
SQ SEQUENCE 354 AA; 38512 MW; B50D7C638CF27091 CRC64;

Query Match 15.4%; Score 332; DB 5; Length 354;
Best Local Similarity 26.2%; Pred. No. 4.9e-20;
Matches 101; Conservative 74; Mismatches 135; Indels 76; Gaps 15;

QY 19 VDNLRGK-----SGGYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFL- 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 13 VTEIRGALGDPIILNTNNMQFYGIIGTIPQSFKLLMDTGSSNFWV-----PSIN 67
QY 65 -----HRYQRLSSTYRDLRKGVYVYQGWEGELGTLVSIPIHGPN-----VTVR 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 CDOSMACRDHAKYDSSKSTFTKSGRIRYSGGVVRGITSIDNVGV--GPATVTQYKF 125
QY 116 AATESDKFFINGSNWGILGLAYAEIARPDSDLEPFDSLVKQTHVPN-LFSLQLCGAG 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 AEMHSDGKLFRAKYDIGIFGLAFPSLSQ--NNQLPFDAMVQGVVRQAVFSLYL--SK 181
QY 175 FPLNQSEVLASVSGSMITIGDHSITGSLWYTPIRREWYEVILVRVEINGODLKM-DC 233
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 QPSEQN-----GGEIYFGINAQRVTGAIHYVPVSOAAHWQVMDNINVGQTTLCVGGC 235
QY 234 KEYNDKSIYDVGTTNLRIPKPKVFEAAVKSIAASSTKEKPDGFWLGEOLVQCAGTTPW 293
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 -----PTVVDVDSGTSLSGSP-----SADVETLNRVIGATKAAGY-----FEVNCATI 277
QY 294 NIFPVISLYLMGE---VTNQSFRTILPQYLRPEVDVATSDQDCYKFAISQSSTGT--- 347
```

```
Db 278 SSLPITNLNGKSPLOQEAYTIRI-----PLTTGGEQCFTRISESASGTNLW 327
QY 348 VNGAVIMEGFVYVDRARKRIGFAVS 373
Db 328 ILGAVFTQTYTVFDRAGNRVGFATA 353

RESULT 14
Q9BGU5 PRELIMINARY; PRT; 386 AA.
AC Q9BGU5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE CATHEPSIN D (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Higuchi M., Miyashita N., Nagamine Y., Awata T.;
RT "Complementary DNA sequence of bovine cathepsin D."
CC Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AB05312; BAB21620.1;
DR HSP; P07339; ILVB.
DR MEROPS; A01.009;
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp.1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Aspartyl protease; Hydrolase.
FT NON_TER 1
SQ SEQUENCE 386 AA; 41965 MW; 33BF00B80585490D CRC64;

Query Match 14.5%; Score 312.5; DB 6; Length 386;
Best Local Similarity 28.0%; Pred. No. 2.5e-18;
Matches 113; Conservative 69; Mismatches 131; Indels 91; Gaps 20;

QY 8 EPG-RRGSFVEMVDNLGRKSGQYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFL-- 64
Db 35 EPAVRQGPPELKLKNYMDAQ---YGEIGIGTTPQCFTVVDGTSANLWVPSIHKLLDI 91
QY 65 ---HRYQRLSSTYRDLRKGVY--VPYQGWEGELGTLVSIPIHGPN-----VTVR 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 ACWTHRYNDKSTY--VKNQTTFDIHGSGSLGSLGSLQSDTVSVPCNPSSSPGGVTQ 149
QY 113 ANI--AAITESDKFFINGSNWGILGLAYAEIARPDSDLEPFDSLVKQTHVPN-LFSLQ 169
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 RQTFGEAIKPGVVFI-AAKEFGILGMAYPRIS--VNNVLPVFDNLMOQLVKNVFS-- 204
QY 170 LCAGAPLNQSEVLASVSGSMITIGDHSITGSLWYTPIRREWYEVILVRVEINGQDL 229
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 -----FFLNR-DPKAPGGELMLGTDSDKYRGLSFHNVTRQAYWQIHMQLDQV-GSSL 257
QY 230 KMDCKEYNDKSIYDVGTTNLRIPKPKVFEAAVKSIAASSTKEKPDGFWLGEOLVQCAG 289
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 TV-CK--GGEALVDGTSLIVGPVEEVLQKAI-----G 290
QY 290 TTPWNIFPVISLYLMGEVNTNQSFRTILPQYLR-PEVDVATSDQDCYKFAISQSSTGT- 347
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 AVP-----LIQGEYMPCEKVSSLQVTVKLGKDYAKSPED-YALKYSQAGTTVC 340
QY 348 -----VNGAVIMEGFVYVDRARKRIGFAVS 374
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 LSGFMGMGMDIPPGGPLWILGDVFIGRYTVFDRDQNRVGLAEAA 384
```

**us-09-724-571-58.rspt**

Search completed: August 7, 2002, 09:17:13  
Job time: 217 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:14:48 ; Search time 22.58 Seconds  
(without alignments)  
440.266 Million cell updates/sec

Title: US-09-724-571-58  
Perfect score: 2156  
Sequence: 1 ETDEEPEPGRSGFVEMVD.....GPFVTLDMEDCGYNIPQTD 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2150	99.7	501	4	US-09-009-191-2
2	2073	96.2	774	4	US-09-009-191-4
3	1132	52.5	518	3	US-08-999-723-2
4	1132	52.5	518	4	US-09-434-427-2
5	1116.5	51.8	514	4	US-09-717-432-2
6	1116.5	51.8	514	4	US-09-912-484-2
7	299.5	13.9	396	1	US-08-208-007A-13
8	299.5	13.9	396	4	US-09-032-523-9
9	295.5	13.7	412	1	US-08-208-007A-12
10	295.5	13.7	412	4	US-08-974-691-4
11	279.5	13.0	458	6	5217891-15
12	273.5	12.7	409	1	US-09-640-305-6
13	273.5	12.7	409	1	US-08-360-673-6
14	273.5	12.7	427	2	US-08-846-021A-8
15	270	12.5	410	1	US-08-088-633-2
16	270	12.5	410	1	US-08-245-756-2
17	270	12.5	410	1	US-08-441-750-2
18	270	12.5	410	2	US-08-441-751-2
19	270	12.5	410	5	PCT-US92-02521-2
20	250	11.6	349	4	US-09-032-523-3
21	227	10.5	398	1	US-08-328-314-2
22	227	10.5	398	1	US-08-731-045-2
23	212	9.8	419	4	US-08-974-691-3
24	210	9.7	397	3	US-09-079-415-2
25	196.5	9.1	430	1	US-08-535-237-2
26	195.5	9.1	427	1	US-07-958-222A-2
27	193	9.0	330	3	US-08-115-753-1

28	193	9.0	419	3	US-08-115-753-2
29	193	9.0	419	3	US-08-115-753-3
30	187	8.7	420	4	US-09-008-271A-4
31	187	8.7	420	4	US-08-974-691-8
32	184.5	8.6	395	1	US-08-723-938-3
33	184.5	8.6	395	2	US-09-080-538-3
34	184	8.5	445	4	US-08-974-691-6
35	184	8.5	451	4	US-08-974-691-2
36	149	6.9	437	4	US-09-353-332-2
37	129.5	6.0	140	3	US-09-211-631-13
38	129.5	6.0	140	4	US-09-285-628-13
39	129.5	6.0	140	4	US-09-001-141-11
40	129.5	6.0	140	4	US-09-532-803-6
41	129.5	6.0	140	4	US-09-653-403-14
42	97	4.5	1030	4	US-09-091-117-2
43	95.5	4.4	280	4	US-09-160-246-14
44	89.5	4.2	1097	2	US-08-680-526-39
45	88	4.1	746	2	US-08-838-219B-6

## ALIGNMENTS

RESULT 1  
US-09-009-191-2  
; Sequence 2, Application US/09009191  
; Patent No. 6319689  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID  
; APPLICANT: CHAPMAN, CONRAD  
; APPLICANT: MURPHY, KAY  
; APPLICANT: SMITH, TRUDI  
; TITLE OF INVENTION: ASP2  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,191  
; FILING DATE: 20-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UK 9701684.4  
; FILING DATE: 28-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23 031  
; REFERENCE/DOCKET NUMBER: GH-70368  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 501 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-009-191-2

Query Match 99.7%; Score 2150; DB 4; Length 501;  
Best Local Similarity 99.8%; Pred. NO. 9.4e-223;







```

; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASp1)
; FILE REFERENCE: GP-70663-D1
; CURRENT APPLICATION NUMBER: US/09/912,484
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/717,432
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 514
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
US-09-912-484-2

Query Match          51.8%; Score 1116.5; DB 4; Length 514;
Best Local Similarity 53.7%; Pred. No. 1.9e-111;
Matches 209; Conservative 66; Mismatches 109; Indels 5; Gaps 3;

QY 8 EPGR-RGSFVEMVNDLRKSGQGYVEMTVGSPPTNLNLVDTGSSNFVAGAAPHPFLHR 66
DB 65 EPVRATANFLAMVNDLQDSGRGYVLEMLICTPPKQVQILVDTGSSNFVAGAPHSYIDT 124
QY 67 YQROLSTYRDRLKGVYVPYTOGKWEGELGDLVSIPIHGNVTVVRANIAAITESDKFFI 126
DB 125 VFDSESSYTHSGKGDVTVKVTQGSWTGFGVEDLVTPKGFNSFLNIATIFSENFEL 184
QY 127 NSGNNEGILGLAYAIARIPDDSLPEFFDSLVKQTHVPNLFSLQLCGAGFPPLNQSEVLASV 186
DB 185 PGIKNNGILGLAYALAKALPSSLETFFDSLVQAQAKIPDIFSMQMGAGLPVAGS---GTN 241
QY 187 GSGMIIGDHSXYTGLSWYPIRREWTYEVILVVEINGODLMDCKEYNDKSIYVDSG 246
DB 242 GCSVLGGIEPLSYKGDWYTPKEWYQIEILKLEIGGONLNDCREYNADRAIVDSG 301
QY 247 TTNLRLPKKVFVAAVKSIAKASSTKFPDGLVGLQVQWAGTTPPWNIFFVISLYLME 306
DB 302 TILLRLPQKVFVAVNAVARTSLIPEFSDGFTWGAQLACWTNSETPWAFYFKISYLRDE 361
QY 307 VTNQSPRITLPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVINMEGFYVDFDARK 366
DB 362 NASRSPRITLPOLYIQPMMGAGFNY-ECYRFGISSTNALVIGATVMEGFYVDFDARQ 420
QY 367 RGFVAVSACHVDFEFTAAVEGPFVTLDM 395
DB 421 RVGFVAVSPCAIEGTTVSEISGPFSTEDI 449

RESULT 7
US-08-208-007A-13
; Sequence 13, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A

; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASp1)
; FILE REFERENCE: GP-70663-D1
; CURRENT APPLICATION NUMBER: US/09/912,484
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/717,432
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 514
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
US-09-912-484-2

Query Match          13.9%; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.9%; Pred. No. 1.2e-23;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 3 DEPEPEPRGRSGFVEMVNDLRKSGQGYVEMTVGSPPTNLNLVDTGSSNFVAGAA--- 58
DB 63 DQSAKEP-----LINVLD-----MEYFGTISIGSPQNFTVIFDTGSSNLWVPSVYCT 110
QY 59 APHPFLHRYQROLSTYRDRLKGVVVPYTOGKWEGELGDLVSIPIHGNVTVVRANIAAI 118
DB 111 SPACKTHSRFPQSOSSTYSQPGOSFSIQYGTGSLGIIGADQVSV-EGLTVVGQGFESV 169
QY 119 TESDKFFINGSNNEGILGLAYAEIARIPDDSLPEFFDSLVKQTHVPNLFSLQLCGAGFP 178
DB 170 TEPQOTFVD-AEFDGILGLQVPSLA--VGGVTPVFONMAQ-----NLVDLPMF SVYSSN 222
QY 179 QSEVLASVSGSMIIGDHSXYTGLSWYPIRREWTYEVILVVEINGODLMDCKEYNY 238
DB 223 PE---GGAGSELGPIGGYDHSFSGSLNWPVTQOAYWQIALDNIQVGG--TVMFCSE--G 275
QY 239 DKSIVDGTTNLRPKKVFVAAVKSIAKASSTKFPDGLVGLQVQWAGTTPWNIFFV 298
DB 276 COALVDGTSLITGSPDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322
QY 299 ISLYLMGEVTVNQSPRITLPQOYLRPVEDVATSDDCYKFAISQSSTG----- 346
DB 323 VTFTING-----VPYTLSPATY--TLLDFVDGMQFC-----SSGFOGLDIHPGAP 366
QY 347 -TVMGAVIMEGFYVDFDARKRIGFA 371
DB 367 LWILGDVFIROFYFVDFRGNRNVGLA 392

RESULT 8
US-09-032-523-9
; Sequence 9, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
```

ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/032,523  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0479 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-855-0555  
 TELEFAX: 650-845-4166  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 396 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: Genbank  
 CLONE: 181994  
 US-09-032-523-9

Sequence 4, Application US/08974691  
Patent No. 6225103  
GENERAL INFORMATION:  
APPLICANT: Keolsch, Gerald  
APPLICANT: Lin, Xinli  
APPLICANT: Tang, Jordan  
TITLE OF INVENTION: Cloning and Characterization of Napsin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 W. Peachtree  
STREET: St.  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974.691  
FILING DATE: 20-NOV-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,196  
FILING DATE: 20-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,126  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMRF 166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-873-8794  
TELEFAX: 404-873-8795  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-974-691-4

Query Match 13.7%; Score 295.5; DB 4; Length 412;  
Best Local Similarity 28.5%; Pred. No. 3.3e-23;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;  
Qy 30 YVEMTVGSPPTLNILVDTGSSNFAVGAAPHPL-----HRYQRLSSYRDLRKGV 83  
Db 79 YGIGTGTPQCTVFTVDGSSNLVPSIHCKLLDIACVHHKYNDSKSSITVKNKTSF 138  
Qy 84 YVPTQCKWEGELGTDLVSP-----HGPNTVVRANIAAITESDKFFINGSNWEGI 134  
Db 139 DIHYGSGSLGYLSQDVTVPQCOSASSALGVKVRQVGEATKPGTITFAAKFDGI 198  
Qy 135 LGAYAEIARPDSDLEFFDSLVKQTHV-PNLSLQLCGAGFPLNOSEVLASVGGSMIIG 193  
Db 199 LGMAYPRIS--VNNVLPVFDNMQKLVDONIFSYL-----SRDPAQPGGELMLG 248  
Qy 194 GIDHSLYTGSLWYPIREWEYEVILVRVEI-NGDCLKMCKEYNDKSIVDSCITNLR 252  
Db 249 GTDSRYKGSLSYLVNTRKAWQVHLDOVEASGLTL---CKE---GCEAIVDTGSLMWG 303  
Qy 253 PKKVFEEAVKSIKAASSTKFPDGFGLGEQLV-CWQAGTTPWNTFPVISLYLMEVTNQS 311  
Db 304 PVDEYRELQKAIGAVPLIQ-----GEYMIPECVKVST-----LPAITLKLGG-----KG 346

Qy 312 FRITLPQOYLRPVEDVATSDDCYKFAISO-----SSTGTVMGAVIMEGFYVVFDRARK 366  
Db 347 YKLS--PEDYTLKVSQAGKTL-CLSGFMGMDIPPSGPLWILGDVFIQRYTYVFDNRN 402  
Qy 367 RIGFAVSA 374  
Db 403 RVGFEEAA 410  
RESULT 11  
5217891-15  
Patent No. 5217891  
APPLICANT: BRAKE, ANTHONY J.; VAN DEN BERG, JOHAN A.  
TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES  
A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS  
POLYPEPTIDES  
NUMBER OF SEQUENCES: 23  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/507,398  
FILING DATE: 09-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 78,551  
FILING DATE: 28-JUL-1987  
SEQ ID NO: 15:  
LENGTH: 458  
5217891-15

Query Match 13.0%; Score 279.5; DB 6; Length 458;  
Best Local Similarity 26.8%; Pred. No. 2.1e-21;  
Matches 95; Conservative 67; Mismatches 128; Indels 65; Gaps 15;  
Qy 31 YVEMTVGSPPTLNILVDTGSSNFAVGAAPHPL-----HRYQRLSSYRDLRKGV 83  
Db 152 YFGIYLGTPQEEFTVLFDTGSSDFW---PSIYCKSNACKNHQRFQDKSSTFQNLGKDL 208  
Qy 84 YVPTQCKWEGELGTDLVSPHGPNTVVRANIAAITESDKFFINGSNWEGILGAYAEIA 143  
Db 209 SIHYGTGSMOGLGYDVTIVSNIVDITQOTVGLSTGEPGVF--TYAEFDGILGMAYPSLA 266  
Qy 144 RPDSLEPFDDSLVKQTHV-PNLSLQLCGAGFPLNQSEVLASVGGSMIIGGTDHSLYTG 202  
Db 267 --SEYISIPVEDNMNRHLVAQDLFSVYMDRNG---QESMLT-----LGAIDPSYTG 313  
Qy 203 SLWYTPTRREWEYEVILVRVEINGDCLKMD--CKEYNDKSIVDSGTTNLRPKKVFEEA 260  
Db 314 SLHWVPVTVOQYVQFTVDSVTISGVVYVACGGC-----QAILDTGTSKLVGSPSSDLNI 367  
Qy 261 VKSIKAASSTKFPDGFGLGE-QLVCWOAGTTPWNIPVVISLYLMGEVTNQSPTITLPO 319  
Db 368 QQAICATQNG-----YGEFDIDCDNLSPYPTVVF-----EINGKMYPLT--PS 408  
Qy 320 QYLRPVEDVATSDDCY---KFAISOSSGTVMGAVIMEGFYVVFDRARKRIGFA 371  
Db 409 AY-----TSQDQGFCTSGFQSENHSGQWILGDVFIQRYTYVFDNRNVLGLA 455

RESULT 12  
US-09-640-305-6  
Sequence 6, Application US/09640305  
Patent No. RE37447  
GENERAL INFORMATION:  
APPLICANT: Fleer, Reinhard  
Fournier, Alain  
Yeh, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville

STATE: PA  
COUNTRY: USA  
ZIP: 19002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 16-Aug-2000  
PRIORITY APPLICATION NUMBER: US/09/640,305  
FILING DATE: 06-FEB-1995  
APPLICATION NUMBER: US/08/360,673  
FILING DATE: 23-JUN-1993  
APPLICATION NUMBER: FR 92/07785  
FILING DATE: 25-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Julie K.  
REGISTRATION NUMBER: 38,619  
REFERENCE/DOCKET NUMBER: ST92040-US  
TELEPHONE: (610)454-3839  
TELEFAX: (610)454-3808  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 409 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-640-305-6

Query Match 12.7%; Score 273.5; DB 1; Length 409;  
Best Local Similarity 25.3%; Pred. No. 7.7e-21;  
Matches 92; Conservative 59; Mismatches 144; Indels 69; Gaps 11;  
QY 30 YVEMTVGSPPTLNILVDTGSSNFAVGAAP----HPFLHRYQRLSSTYRDLRKGVV 85  
DB 96 YFEITLGPSPQSFVILDTGSSNLWVPSAECGLACFLHTKYDHEASSTYKANGSEFAI 155  
QY 86 PYTQKWEGLGTDLVSPHGNVTVRANIAITAESDRKFFINGSNWEGILGLAYAEIARP 145  
DB 156 QYSGSLEGVYGRDLTI--GDLVTPDQDEATSEPGLAFAFGKFDGILGLAYDSIS-V 212  
QY 146 DDLSEPPFDSLKQTHVNLFSQLCGAGFPLNOSSEVLASVGGSMIIGIDHSLYTGSLW 205  
DB 213 NRIVPPVNAIKNLDDPFVA-----FYLGDSD-KSEDDGEASFGGIDEKYTGTEIT 264  
QY 206 YTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDVGTTNLRPLPKVFEAAVKSIIK 265  
DB 265 WLPVRRKAYWE-----VKFEGIGLEEYATLEGHGAIDTGTSLIALPSGLAETLNAEIG 319  
QY 266 AASSTEKEPDPGFWLGEQVLCVQAGTTPNWIPVVISLYLMGEVNTNQSFRIT----- 315  
DB 320 AKKG-----WSGOYSVDCESRDS-----LPDLTLNFG-----YNFTIAYDTYLEVSG 363  
QY 316 -----ILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGF 370  
DB 364 SCISAFTPMDFPEVGPGLA-----IIGDAFLRKYSIYDIGHDAVGL 405  
QY 371 AVSA 374  
DB 406 AKAA 409

RESULT 13  
US-08-360-673-6  
Sequence 6, Application US/08360673  
Patent No. 5679544  
GENERAL INFORMATION:

APPLICANT: Fleer, Reinhard  
APPLICANT: Fournier, Alain  
APPLICANT: Yeh, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/360,673  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR93/00623  
FILING DATE: 23-JUN-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 92/07785  
FILING DATE: 25-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Julie K.  
REGISTRATION NUMBER: 38,619  
REFERENCE/DOCKET NUMBER: ST92040-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)454-3839  
TELEFAX: (610)454-3808  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 409 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-360-673-6

Query Match 12.7%; Score 273.5; DB 1; Length 409;  
Best Local Similarity 25.3%; Pred. No. 7.7e-21;  
Matches 92; Conservative 59; Mismatches 144; Indels 69; Gaps 11;  
QY 30 YVEMTVGSPPTLNILVDTGSSNFAVGAAP----HPFLHRYQRLSSTYRDLRKGVV 85  
DB 96 YFEITLGPSPQSFVILDTGSSNLWVPSAECGLACFLHTKYDHEASSTYKANGSEFAI 155  
QY 86 PYTQKWEGLGTDLVSPHGNVTVRANIAITAESDRKFFINGSNWEGILGLAYAEIARP 145  
DB 156 QYSGSLEGVYGRDLTI--GDLVTPDQDEATSEPGLAFAFGKFDGILGLAYDSIS-V 212  
QY 146 DDLSEPPFDSLKQTHVNLFSQLCGAGFPLNOSSEVLASVGGSMIIGIDHSLYTGSLW 205  
DB 213 NRIVPPVNAIKNLDDPFVA-----FYLGDSD-KSEDDGEASFGGIDEKYTGTEIT 264  
QY 206 YTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDVGTTNLRPLPKVFEAAVKSIIK 265  
DB 265 WLPVRRKAYWE-----VKFEGIGLEEYATLEGHGAIDTGTSLIALPSGLAETLNAEIG 319  
QY 266 AASSTEKEPDPGFWLGEQVLCVQAGTTPNWIPVVISLYLMGEVNTNQSFRIT----- 315  
DB 320 AKKG-----WSGOYSVDCESRDS-----LPDLTLNFG-----YNFTIAYDTYLEVSG 363  
QY 316 -----ILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVVDRAKRIGF 370  
DB 364 SCISAFTPMDFPEVGPGLA-----IIGDAFLRKYSIYDIGHDAVGL 405  
QY 371 AVSA 374

QY	30	YVENVTCGSPQTLNLTLDVTGSSNFAVG----	APHPEFLHRYQORQSSDLSYRILKARAGV	155
DB	96	YFTEVSLGTTPQSPKVIYLDTGSSNLWVP	KDCGSLACFLHAKYDHDESSYTKKNGSF	145
QY	86	PYTQCKWEGELGTLDSIPIHGPMVTVR	ANIAITESDKFFINGSNWBGILGLAYAIAR	145
DB	156	RYGSGSMGYYSQDVLIQ----	GDUTIPKVDFAEATSEGPLAFACFKDGI	207
QY	146	DDSLPEPFFDSLVKQTHVPNLF----	SLQLCGA----	199
DB	208	-----DSISVNKIVPPIYKALEL	LDLDEPKFAFYLGTDKDESDGLATFG	259
QY	200	YNGSLWYTPIRREYYEVIIVRVEING	ODLKMDCKEYNNDKSIYDVGSTG	259
DB	260	YESKITWLPVRKAYWE-----	VSFDGVGLGSEYAELOQTGA	314
QY	260	AVKSIKAASSTKEKPDGFWLGEOLVC	WQAOGTTTWNIFPVISLYLMGEVNT	319
DB	315	LNAEIGATKG-----	WSQGVAVDCDTRDS-----	356
QY	320	QYLRPEVDVATSDDCYKFAISQSTGT	VMAIMEGEYVWFDRARKRIGFAVS	373

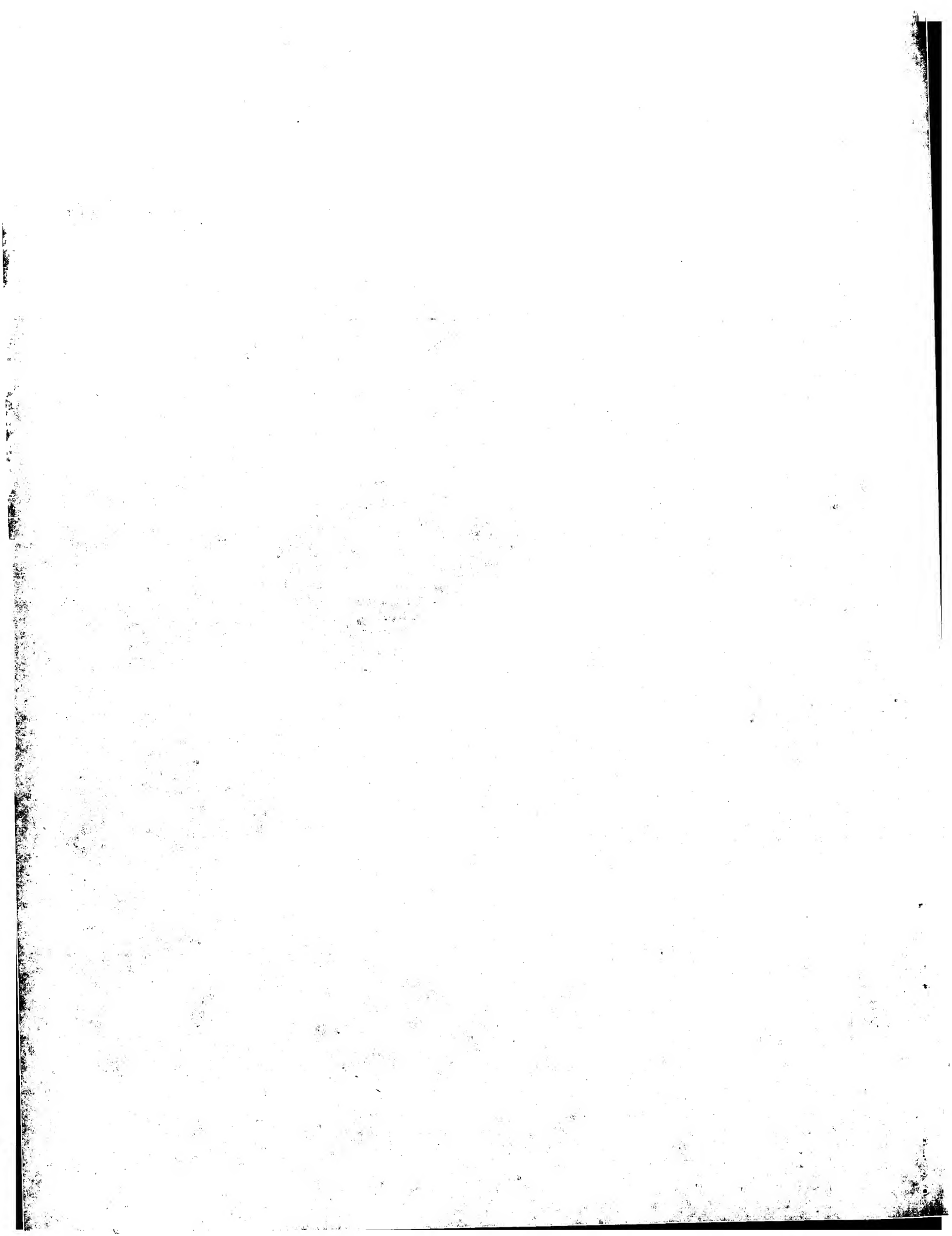
Wed Aug 7 10:38:05 2002

us-09-724-571-58.rai

Page 9

Db 357 DYTLEVSSCISAFPMDFP-EPIGPLAIIIGDSFLRKYYVYD LGKDAVGLAKS 409

Search completed: August 7, 2002, 09:14:49  
Job time: 74 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:35 ; Search time 56.55 Seconds  
(without alignments)  
895.661 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEPPEPGRGSFVEMV.....CLRLRQHDDFADDISLLK 456

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2419	100.0	456	21	AA07897
2	2419	100.0	488	21	AA07897
3	2419	100.0	488	22	AA07897
4	2419	100.0	501	21	AA07897
5	2419	100.0	501	21	AA07897
6	2419	100.0	503	22	AA07897
7	2419	100.0	503	22	AA07897
8	2414	99.8	501	21	AA07897
9	2414	99.8	501	22	AA07897
10	2414	99.8	501	22	AA07897
11	2414	99.8	501	22	AA07897

12	2414	99.8	501	22	AAU07202	Human aspartyl pro
13	2414	99.8	501	22	AAE02581	Human aspartyl pro
14	2413	99.8	501	19	AAW59807	Amino acid sequenc
15	2392	98.9	501	21	AA074769	Rat beta-secretase
16	2390	98.8	501	21	AA074768	Murine beta-secret
17	2390	98.8	501	21	AA074768	Murine aspartyl pr
18	2390	98.8	501	22	AAE10631	Murine aspartyl pr
19	2390	98.8	501	22	AAE06861	Murine aspartyl pr
20	2390	98.8	501	22	AAU06605	Mouse aspartyl pr
21	2390	98.8	501	22	AAU07204	Mouse aspartyl pr
22	2390	98.8	501	22	AA074948	Mouse aspartyl sec
23	2390	98.8	501	22	AAE02583	Murine aspartyl pr
24	2351.5	97.2	969	22	ABG09611	Novel human diagno
25	2320	95.9	790	19	AAW59808	Partial amino acid
26	2264.5	93.6	476	21	AA0788426	Human aspartyl pro
27	2264.5	93.6	476	21	AA0788426	Human aspartyl pro
28	2264.5	93.6	476	22	AAE10630	Human aspartyl pro
29	2264.5	93.6	476	22	AAE06860	Human aspartyl pro
30	2264.5	93.6	476	22	AAU06604	Human aspartyl pro
31	2264.5	93.6	476	22	AAU07203	Human aspartyl pro
32	2238.5	92.5	476	22	AAE02582	Human aspartyl pro
33	2238.5	92.5	476	22	AAE06909	Human aspartyl pro
34	2212	91.4	427	22	AA02619	Murine aspartyl pr
35	2160	89.3	453	21	AA0788438	Murine aspartyl pr
36	2160	89.3	453	21	AA0788438	Human polypeptide,
37	2160	89.3	453	22	AAE10642	Modified human asp
38	2160	89.3	453	22	AAE06872	Human-Asp 2(a) pro
39	2160	89.3	453	22	AAU06616	Human-Asp 2(a) del
40	2160	89.3	453	22	AAU07215	Human-Asp 2(a) del
41	2160	89.3	459	21	AA0788439	Human-Asp 2(a) del
42	2160	89.3	459	22	AAE10643	Modified human asp
43	2160	89.3	459	22	AAE06873	Human-Asp 2(a) pro
44	2160	89.3	459	22	AAU06617	Human-Asp 2(a) del
45	2160	89.3	459	22	AAU07216	Human aspartyl pro

## ALIGNMENTS

RESULT 1

AA07897

ID AAB07897 standard; Protein; 456 AA.

XX AC

XX AAB07897;

XX DT 14-NOV-2000 (first entry)

XX DE Active enzyme portion of human beta-secretase enzyme.

XX KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;

XX KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

XX KW inhibitor; ss.

XX OS Homo sapiens.

XX PN WO200047618-A2.

XX PD 17-AUG-2000.

XX PF 10-FEB-2000; 2000WO-US03819.

XX PR 10-FEB-1999; 99US-0119571.

XX PR 15-JUN-1999; 99US-0139172.

XX PA (ELAN-) ELAN PHARM INC.

XX PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

XX PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX WPI; 2000-533011/48.

XX DR Purified beta-secretase protein used in assays to discover inhibitors

XX PT which can be used for the treatment of amyloidogenic diseases e.g.

```
PT Alzheimer's disease
XX
XX Claim 24; Fig 2B; 121pp; English.
XX
XX The specification describes a beta-secretase enzyme. The enzyme cleaves
XX beta-amyloid precursor protein to produce beta-amyloid peptide. This
XX enzyme is therefore implicated in the production of amyloid plaque
XX components which accumulate in the brains of individuals afflicted with
XX Alzheimer's disease. Inhibitors of beta-secretase are administered to
XX a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
XX disease-like pathology to test if they maintain or improve cognitive
XX ability or reduce the plaque burden. The compounds are used for the
XX treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
XX present sequence represents the active enzyme portion of human
XX beta-secretase enzyme.
XX Sequence 456 AA;
SQ
Query Match 100.0%; Score 2419; DB 21; Length 456;
Best Local Similarity 100.0%; Pred. No. 3.9e-242;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEPEGRGSGFVEMVDNLKSGQGYVEMTVGSPQTNLILVDTGSSNFVAGAAP 60
DB 1 etdeepeeegrgrsgfvmvndnlrgskggyvemtvgspqtnilvdtgssnfavgaa 60
QY 61 HPFLHRYQRLSTYRDLRGVYVPTQGWEGELGDLVSIHPGPNVTVRANIAAITE 120
DB 61 hpflhryqrqlsstyrdlrgvvyvptqgkwegelgdlvsihpgpnvtvranaia 120
QY 121 SDKFFINGSNWEGILGLAYAEARTDDSLPEFDSLVKQTHVPLNLSLQLCGAGFPLNQS 180
DB 121 sdkffingsnwegilglayaeartddslpefdslvkqthvplnlsfqlcgagfpln 180
QY 181 EYLASVGGSMIIGGDHSLYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNYDK 240
DB 181 eylvavsgsmiiggdhsltytgslytswytpirrewyvevliivrveingqdlkmdck 240
QY 241 SIVDSGTTNLRPKKVFEEAIVKSIKAASTTEKFPDGFVWLGEOVCWQAGTTPWNIFFVIS 300
DB 241 sivdsgettlnrlpkkvfeavksikaasttekefpdgfvwlgvewlgeqlvcwagtpw 300
QY 301 LYLMGEVNTNQSPRITILPQQYLIRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
DB 301 lylmgevntnqspfrtilpqylirpvedvatsddcykfaissstgtvmgavimegf 360
QY 361 FDRARRKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 420
DB 361 fdrarrkrigfavsachvhdefrtaaavegpfvtldmedcgynipqdestlmtiayv 420
QY 421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
DB 421 calfmplclmvcwqrclrclrqhghddfaddisllk 456
RESULT 2
AAB66572
ID AAB66572 standard; Protein; 488 AA.
XX
XX AAB66572;
XX
XX 12-APR-2001 (first entry)
XX
XX Human memapsin 2.
XX
XX Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;
XX APP; memapsin 2 inhibitor; Alzheimer's disease.
XX
XX Homo sapiens.
XX
XX WO200100665-A2.
XX
```

```
PD 04-JAN-2001.
XX
XX 27-JUN-2000; 2000WO-US17742.
XX
XX 28-JUN-1999; 99US-0141363.
XX 30-NOV-1999; 99US-0168080.
XX 25-JAN-2000; 2000US-0177836.
XX 27-JAN-2000; 2000US-0178368.
XX 08-JUN-2000; 2000US-0210292.
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX (UNII ) UNIV ILLINOIS FOUND.
XX
XX Tang JJN, Hong L, Ghosh AK;
XX
XX WPI; 2001-137933/14.
XX N-PSDB; AAF31848.
XX
XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2
XX having 2 catalytic aspartic residues and substrate binding cleft, used
XX to treat Alzheimer's disease by blocking amyloid precursor protein
XX cleavage
XX
XX Example 1; Page 72-74; 86pp; English.
XX
XX The present sequence is given in a specification relating to an inhibitor
XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
XX active site, which is defined by the presence of two catalytic aspartic
XX residues and a substrate binding cleft. The inhibitor is useful for
XX the treatment and diagnosis of Alzheimer's disease. It is useful in
XX screens for individuals with a genetic predisposition to Alzheimer's
XX disease. The inhibitor is useful as a reagent for specifically binding to
XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
XX isolation, purification and characterisation.
XX
XX Sequence 488 AA;
SQ
Query Match 100.0%; Score 2419; DB 22; Length 488;
Best Local Similarity 100.0%; Pred. No. 4.3e-242;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEPEPEGRGSGFVEMVDNLKSGQGYVEMTVGSPQTNLILVDTGSSNFVAGAAP 60
DB 33 etdeepeeegrgrsgfvmvndnlrgskggyvemtvgspqtnilvdtgssnfavgaa 92
QY 61 HPFLHRYQRLSTYRDLRGVYVPTQGWEGELGDLVSIHPGPNVTVRANIAAITE 120
DB 93 hpflhryqrqlsstyrdlrgvvyvptqgkwegelgdlvsihpgpnvtvranaia 152
QY 121 SDKFFINGSNWEGILGLAYAEARTDDSLPEFDSLVKQTHVPLNLSLQLCGAGFPLNQS 180
DB 153 sdkffingsnwegilglayaeartddslpefdslvkqthvplnlsfqlcgagfpln 212
QY 181 EYLASVGGSMIIGGDHSLYTGSLWYTPIRREWYVEVLIIVRVEINGQDLKMDCKEYNYDK 240
DB 213 eylvavsgsmiiggdhsltytgslytswytpirrewyvevliivrveingqdlkmdck 272
QY 241 SIVDSGTTNLRPKKVFEEAIVKSIKAASTTEKFPDGFVWLGEOVCWQAGTTPWNIFFVIS 300
DB 273 sivdsgettlnrlpkkvfeavksikaasttekefpdgfvwlgvewlgeqlvcwagtpw 332
QY 301 LYLMGEVNTNQSPRITILPQQYLIRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
DB 333 lylmgevntnqspfrtilpqylirpvedvatsddcykfaissstgtvmgavimegf 392
QY 361 FDRARRKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 420
DB 393 fdrarrkrigfavsachvhdefrtaaavegpfvtldmedcgynipqdestlmtiayv 452
QY 421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
DB 453 calfmplclmvcwqrclrclrqhghddfaddisllk 488
```

RESULT 3  
AAB61334  
ID AAB61334 standard; protein; 488 AA.  
XX  
AC AAB61334;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Memapsin 2 protein.  
XX  
KW Memapsin 2; catalyst; Alzheimer's.  
XX  
OS Homo sapiens.  
XX  
PN WO200100663-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-US17661.  
XX  
PR 28-JUN-1999; 99US-0141363.  
PR 30-NOV-1999; 99US-0168080.  
PR 25-JAN-2000; 2000US-0177836.  
PR 27-JAN-2000; 2000US-0178368.  
PR 08-JUN-2000; 2000US-0210292.  
XX  
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Tang JJN, Lin X, Koelsch G;  
XX  
DR WPI; 2001-102885/11.  
XX  
PT Purified recombinant catalytically active memapsin 2, used to screen  
PT inhibitors of it, which are used to treat and prevent Alzheimer's  
PT disease -  
XX  
PS Claim 2; Page 73-75; 86pp; English.  
XX  
CC The present invention relates to a purified recombinant  
CC catalytically active memapsin 2. The invention may be used for  
CC isolating inhibitors which are used to treat or prevent  
CC Alzheimer's disease. The invention may also be used to screen  
CC for individuals more genetically prone to develop Alzheimer's  
CC disease.  
XX  
SQ Sequence 488 AA;

Query Match 100.0%; Score 2419; DB 22; Length 488;  
Best Local Similarity 100.0%; Pred. No. 4.3e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETDEPEPEGRGSEFVEMVNLKSGQGYVEMTVGSPPTLNLVDTGSSNFVGAAP 60  
DB 33 etdeepeepgrgsfvmvndlrgsggyvemtvgspptlnlvdtgssnfavgap 92  
QY 61 HPFLHRYQRLSTYRDLRKGVVYPTQGWGELCTDLVSPHGNVTVRANIAITE 120  
DB 93 hpfhrryqrqlstydrlrkgyvpytqgweelgtdlvsiphgnvtraniaaite 152  
QY 121 SDRFFINGSNWEGILGAYAEIARPDLSLFFDLSLVKQTHVNLFLSLQLCGAGFPLNQS 180  
DB 153 sdrffingsnwegilgayaearpdslsleffdsllvkthvnlflslqlcgagfplnqs 212  
QY 181 EVLASVGSMLIGDHSILTSGSLWYTPIRREYVYVIVRVEINGDQLKMDCKEYNYDK 240  
DB 213 evlasvgssmligdhsltytsgslwytpirreyvylvrveingdqlkmdckeynydk 272  
QY 241 SIVDSGTTNLRPKKVEAAVKSTKFAASSTKEPKPDGFWLGEQLVCWAGTTPWNIFPVIS 300  
DB 273 sivdsgttnlrpkkveaavksikaasstekipdpgfwlgeqlvcwagttppwnifpvls 332

QY 301 LYLMEVTNQSFRTITLPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
DB 333 lylmevtngsfritilpqgylrpvedvatsgddcykfaissstgtvmgavimegfyv 392  
QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLDMEDCGYNIPOTDESTLMTIAYVMAAI 420  
DB 393 fdrarkrigfavsachvhdeftraavegpfvtldmedcgynlpqdestclmtiayvmaai 452  
QY 421 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLILK 456  
DB 453 calfmlplclmvcqwrcrlrqhghddfaddisllk 488  
RESULT 4  
AAY94767  
ID AAY94767 standard; Protein; 501 AA.  
XX  
AC AAY94767;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human beta-secretase amino acid sequence.  
XX  
KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; human;  
KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..45  
FT /label= putative signal peptide  
FT Protein 46..501  
FT /label= Beta-secretase  
XX  
PN WO200058479-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 23-MAR-2000; 2000WO-US07755.  
XX  
PR 26-MAR-1999; 99US-0277229.  
XX  
PA (AMGE-) AMGEN INC.  
XX  
PI Citron M, Vassar RJ, Bennett BD;  
XX  
DR WPI; 2000-594643/56.  
DR N-PSDB; AAA28278.  
XX  
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful  
PT for diagnosis and gene therapy of Alzheimer's disease -  
XX  
PS Claim 1; Fig 4; 145pp; English.  
XX  
CC This invention relates to 3 nucleotide sequences encoding beta-secretase  
CC proteins. Beta-secretase is an enzyme involved in the production of one  
CC of the components of amyloid plaques involved in the production of one  
CC invention includes an expression vector comprising the nucleotide  
CC sequence, a host cell comprising the expression vector, and a process for  
CC producing the protein through culturing the transformed cells. Also  
CC included in the invention are a polypeptide derivative of the  
CC beta-secretase protein, a fusion protein comprising beta-secretase fused  
CC to a heterologous amino acid sequence, and a method for modulating the  
CC levels of beta-secretase polypeptide in a mammal comprising administering  
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and  
CC nootropic activity. The beta-secretase nucleotide sequence may be used to  
CC map locations of the beta-secretase gene and related genes on chromosomes  
CC and as hybridization probes in diagnostic assays to test for the presence  
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's  
CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be  
CC used as anti-sense inhibitors of beta-secretase expression, in gene  
CC therapy of Alzheimer's disease, and for the identification of compounds

CC that modulate beta-secretase activity. Antibodies to the beta-secretase  
CC protein may be used for in vitro and in vivo diagnostic purposes to  
CC detect the presence of beta-secretase polypeptide in a body fluid or cell  
CC sample. The present sequence represents the human beta-secretase protein.  
XX  
SQ Sequence 501 AA;

Query Match 100.0%; Score 2419; DB 21; Length 501;  
Best Local Similarity 100.0%; Pred. No. 4.5e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETDEPEPEGRGRSFVEMVDNLKCKSGQYVYVMTGSPPTNLILVDTGSSNFVCAAP 60  
Db 46 etdeepeeprrgsfvmvndlrgksggyvmtvgsppqtnilvdtgssnfavgaap 105  
QY 61 HPFLHRYQRLSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGNVTVRANIAAITE 120  
Db 106 hpflhryyqrqlstydrlrkgyvpytqgkwegeltdlvsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLVSKQTHVPLNLSLQLCGAGFPLNQS 180  
Db 166 sdkffingsnwegilglayaeiarpdslslepfdslnkqthvplnlsfqlcagagfplnqs 225  
QY 181 EVLASVGSMTIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNDK 240  
Db 226 evlasvgmsmliggidhsltytgslywtpirrewyevliivrveingodlkmckeynydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGELVLCVQWAGTTPWNIFPVIS 300  
Db 286 sivdsgttnlrpkkvfeavksikaasstekfpgdglvclwqagtpwnifpvls 345  
QY 301 LYLMGEVTNOSFRITILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 346 lylmgevtnosfrilitpqqylrpvedvatsqddcykfaissqstgtvmgavimegyfvv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgytnipqtdstlmtiayvmaai 465  
QY 421 CALFMLPLCLMVCOWRCLRLRQHQHDFADDISLLK 456  
Db 466 calfmlplclmvcowrcrlrlrqhqhdffaddisllk 501

RESULT 5  
AAB07896  
ID AAB07896 standard; Protein; 501 AA.  
AC AAB07896;  
XX  
XX  
XX Amino acid sequence of a human beta-secretase enzyme.  
DE  
XX  
XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200047618-A2.  
FN  
XX  
PD 17-AUG-2000.  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
XX WPI; 2000-533011/48.  
DR N-PSDB; AAA59550, AAA59551.  
XX  
XX Purified beta-secretase protein used in assays to discover inhibitors  
XX which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease  
PT  
XX Claim 17; Fig 2A; 121pp; English.  
XX  
XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme.  
XX  
SQ Sequence 501 AA;

Query Match 100.0%; Score 2419; DB 21; Length 501;  
Best Local Similarity 100.0%; Pred. No. 4.5e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETDEPEPEGRGRSFVEMVDNLKCKSGQYVYVMTGSPPTNLILVDTGSSNFVCAAP 60  
Db 46 etdeepeeprrgsfvmvndlrgksggyvmtvgsppqtnilvdtgssnfavgaap 105  
QY 61 HPFLHRYQRLSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGNVTVRANIAAITE 120  
Db 106 hpflhryyqrqlstydrlrkgyvpytqgkwegeltdlvsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLVSKQTHVPLNLSLQLCGAGFPLNQS 180  
Db 166 sdkffingsnwegilglayaeiarpdslslepfdslnkqthvplnlsfqlcagagfplnqs 225  
QY 181 EVLASVGSMTIGGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGODLKMCKEYNDK 240  
Db 226 evlasvgmsmliggidhsltytgslywtpirrewyevliivrveingodlkmckeynydk 285  
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFGLGELVLCVQWAGTTPWNIFPVIS 300  
Db 286 sivdsgttnlrpkkvfeavksikaasstekfpgdglvclwqagtpwnifpvls 345  
QY 301 LYLMGEVTNOSFRITILPQQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360  
Db 346 lylmgevtnosfrilitpqqylrpvedvatsqddcykfaissqstgtvmgavimegyfvv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgytnipqtdstlmtiayvmaai 465  
QY 421 CALFMLPLCLMVCOWRCLRLRQHQHDFADDISLLK 456  
Db 466 calfmlplclmvcowrcrlrlrqhqhdffaddisllk 501

RESULT 6  
AAB66573  
ID AAB66573 standard; Protein; 503 AA.  
XX  
XX AAB66573;  
AC  
XX  
XX 12-APR-2001 (first entry)  
DT  
XX  
XX Human pro-memapsin 2.  
DE  
XX  
KW Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;

KW APP; memapsin 2 inhibitor; Alzheimer's disease; ss.

OS Homo sapiens.

PN WO200100665-A2.

XX 04-JAN-2001.

XX 27-JUN-2000; 2000WO-US17742.

XX 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PA (UNII ) UNIV ILLINOIS FOUND.

PI Tang JJN, Hong L, Ghosh AK;

XX WPI; 2001-137933/14.

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2

PT having 2 catalytic aspartic residues and substrate binding cleft, used

PT to treat Alzheimer's disease by blocking amyloid precursor protein

PT cleavage -

XX Example 4; Fig 1; 86pp; English.

XX The present sequence is given in a specification relating to an inhibitor

CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2

CC active site, which is defined by the presence of two catalytic aspartic

CC residues and a substrate binding cleft. The inhibitor is useful for

CC the treatment and diagnosis of Alzheimer's disease. It is useful in

CC screens for individuals with a genetic predisposition to Alzheimer's

CC disease. The inhibitor is useful as a reagent for specifically binding to

CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2

CC isolation, purification and characterisation.

XX Sequence 503 AA;

Query Match 100.0%; Score 2419; DB 22; Length 503;

Best Local Similarity 100.0%; Pred. No. 4.6e-242;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGFSFVEMVDNLKRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAP 60

Db 48 etdeepeepgrgrgsfvmvndnlrgksggyvemtvgspqtnilvdtgssnfavgaap 107

QY 61 HPFLHRYYQRLSSTYRDLRKGVVVPYTGQKWEGLGTDLSVIPHGNVTVRANIAAITE 120

Db 108 hpflhryyqrlsstyrdlrkgyvpytgqkwegeigtdivsphgpnvtvranaiaite 167

QY 121 SDRFFINGSNWEGILGLAYAEIARPDSDLVVKQTHVPNLFLSLQLCGAGFPLNQS 180

Db 168 sdrffingsnwegilglayaeiarpddslvfkthvnpnlfsqlcgcagfplnqs 227

QY 181 EVLASVGSMIIGIDHSILTYGSLWYPIRREWYEVIIIVRVEINGODLKMDCKEYNYDK 240

Db 228 evlasvgssmli9gidhslytgslywpcirrewyevliivrveingodlkmckeynydk 287

QY 241 SIVDSGTTNLRPKKVFEEAAVKSAASSTEKFPDGFGLGQVLCVQAGTTPWNIFPVIS 300

Db 288 sivdsghtnlrlpkkvfeaaavksikaasstekfpdgfwlgeqlvcwagttwnifpvvis 347

QY 301 LYLMEVTNSFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGYVYV 360

Db 348 lylmgevtngsfrtilpqqylrpvedvatsqddcykfaissgstgtvmgavimegyfvyv 407

QY 361 FDRARKRIGFAVSACHVHDEFRAAABGPFVTLDMEDCGYNIPQTDSTLMTAYVMAAI 420

|||||

Db 408 fdrarkrigfavsvachvhdefrtaavegpfvtldmedcgynipqtdstlmtayvmaai 467

QY 421 CALEWMLPLCLMVCQWRCLRLRQHQHDFADDISILK 456

Db 468 calfmplclmvcqwrclrlrqhnddfaddisilk 503

RESULT 7

AAB61335

ID AAB61335 standard; protein; 503 AA.

XX

AC AAB61335;

DT 02-APR-2001 (first entry)

XX T7 promoter and vector sequence.

XX Memapsin 2; catalyst; Alzheimer's.

XX Homo sapiens.

OS Synthetic.

XX WO200100663-A2.

XX 04-JAN-2001.

XX 27-JUN-2000; 2000WO-US17661.

XX 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Tang JJN, Lin X, Koelsch G;

XX WPI; 2001-102885/11.

XX Purified recombinant catalytically active memapsin 2, used to screen

PT inhibitors of it, which are used to treat and prevent Alzheimer's

PT disease -

XX Disclosure; Fig 1; 86pp; English.

XX The present invention relates to a purified recombinant

CC catalytically active memapsin 2. The invention may be used for

CC isolating inhibitors which are used to treat or prevent

CC Alzheimer's disease. The invention may also be used to screen

CC for individuals more genetically prone to develop Alzheimer's

CC disease.

XX Sequence 503 AA;

Query Match 100.0%; Score 2419; DB 22; Length 503;

Best Local Similarity 100.0%; Pred. No. 4.6e-242;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGFSFVEMVDNLKRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAP 60

Db 48 etdeepeepgrgrgsfvmvndnlrgksggyvemtvgspqtnilvdtgssnfavgaap 107

QY 61 HPFLHRYYQRLSSTYRDLRKGVVVPYTGQKWEGLGTDLSVIPHGNVTVRANIAAITE 120

Db 108 hpflhryyqrlsstyrdlrkgyvpytgqkwegeigtdivsphgpnvtvranaiaite 167

QY 121 SDRFFINGSNWEGILGLAYAEIARPDSDLVVKQTHVPNLFLSLQLCGAGFPLNQS 180

Db 168 sdrffingsnwegilglayaeiarpddslvfkthvnpnlfsqlcgcagfplnqs 227

QY 181 EVLASVGSMIIGIDHSILTYGSLWYPIRREWYEVIIIVRVEINGODLKMDCKEYNYDK 240

```
|||||
Db 228 evlasvgsmiigdhsltytsglwytpirrewyevilvrveingqdlkmdckeynydk 287
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNIFPVIS 300
|||
Db 288 sivdsgttnlrpkkvfeavksikaasstekfpgdfwlgelvcwgagttppwnifpvls 347
QY 301 LYLMGEVTNOSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVV 360
|||
Db 348 lyImgevtnsgfrtllpqgylrpvedvatsqddcykfaIsqsgstgtvmgavImegfyvv 407
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420
|||
Db 408 fdrarkrigfavsachvhdefrtaavegpfvtdmedcgynipqtdestlmtiayvmaai 467
QY 421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
|||
Db 468 calfmplclmvcqwrclrcrlrgqhddfadddisllk 503

RESULT 8
AA188425
ID AAY88425 standard; Protein; 501 AA.
AC
XX
XX
XX
DT 03-AUG-2000 (first entry)
XX
XX Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
DE
KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
XX Homo sapiens.
OS
XX WO200017369-A2.
PN
XX
XX 30-MAR-2000.
XX
XX 23-SEP-1999; 99WO-US20881.
XX
XX 24-SEP-1998; 98US-0101594.
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
DR N-PSDB; AAA15662.
XX
XX New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Claim 48; Fig 2; 183pp; English.
XX
XX This sequence represents the human aspartyl protease 2 (Asp2) amino acid
XX sequence. The invention relates to a protease (e.g. Asp2) capable of
XX cleaving the beta secretase site of amyloid precursor protein (APP). The
XX protease contains a sequence encoding the amino acid sequence DTG and a
XX sequence encoding DSG or DTG separated by 100-300 amino acids. When
XX mutated the APP gene causes an autosomal dominant form of Alzheimer's
XX disease. APP localises to the cell surface membrane and have a single
XX C-terminal transmembrane domain. Proteolytic processing of APP produces
XX the amyloid beta protein, which is possibly very important in Alzheimer's
XX disease. The invention includes a nucleotide sequence encoding the
XX protease, a vector containing the nucleotide sequence, and a cell line
XX comprising the vector. Methods for screening for inhibitors of beta
XX secretase activity are also given in the invention. The human aspartase
XX protein and nucleotide sequences and the methods for identifying
XX inhibitors of the protease, are useful in the treatment of and research
XX in to Alzheimer's disease.
XX
```

```

SQ Sequence 501 AA;
Query Match 99.8%; Score 2414; DB 21; Length 501;
Best Local Similarity 99.8%; Pred. No. 1.5e-241;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ETDEEPEEPGRGSRFSFVEMVDNLRGKSGQGYVEMTGVSPQPTLNILVDTGSSFAVGAAP 60
|||
Db 46 etdeeepepprrgsfvmvdnlrgksggyvemtgvspptlnilvdtgssnfavgaap 105
|||
QY 61 HPFLHRYQRQLSSTYRDLRKGVYVPYTOCKWEGELGTDLVSPHPGNVTVVRANIAAITE 120
|||
Db 106 hpfLhryyqrqlsstyrdlrkgyvypvtgkwegelgtdivspghpntvtraniaaite 165
|||
QY 121 SDKFFFTNGSNWEGITGLAYAEIARPDSDLPEFFDSLKQTHVPLNLSQLCGAGFPLNQS 180
|||
Db 166 sdkffftngsnwegitglayaeiarpddsllepffdsLvkqthvplnlshlhcsggfpInqs 225
|||
QY 181 EVLASVGSMIIGGIDHSLYTGSLWYTPIRREWYEVIIVRVEINGODLKMCKEYNYDK 240
|||
Db 226 evlasvggsmiigdhsltytsglwytpirrewyevilvrveingqdlkmdckeynydk 285
|||
QY 241 SIYDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNIFPVIS 300
|||
Db 286 sivdsgttnlrpkkvfeavksikaasstekfpgdfwlgelvcwgagttppwnifpvls 345
|||
QY 301 LYLMGEVTNOSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVV 360
|||
Db 346 lyImgevtnsgfrtllpqgylrpvedvatsqddcykfaIsqsgstgtvmgavImegfyvv 405
|||
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420
|||
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtdmedcgynipqtdestlmtiayvmaai 465
|||
QY 421 CALFMLPLCLMVCQWRCLRCLRQHQHDDFADDISLLK 456
|||
Db 466 calfmplclmvcqwrclrcrlrgqhddfadddisllk 501

RESULT 9
AAE10629
ID AAE10629 standard; Protein; 501 AA.
XX
XX
AC AAE10629;
XX
XX
DT 10-DEC-2001 (first entry)
XX
XX Human aspartyl protease 2(a) [hu-Asp2(a)] protein.
DE
XX
XX Human; aspartyl protease 2(a); Asp2(a); amyloid precursor protein; APP;
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
KW amyloid plaque; neuronal loss; proteolytic; nontropic; neuroprotective;
KW chromosome 11q23.3-24.1.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..21
FT Peptide /label= Signal_peptide
FT Peptide 22..45
FT Peptide /label= Asp_2a_prepropeptide
FT Peptide 46..57
FT Peptide /label= Asp_2a_propeptide
FT Protein 58..501
FT Protein /label= Mature_human_Asp_2a_protein
FT Region 420..454
FT Domain /label= Alpha-helical_spacer_region
FT Domain 455..477
FT Domain /label= Transmembrane_domain
FT Domain 478..501
FT Domain /label= Cytoplasmic_domain
XX
```

PN GB2357767-A.  
XX 04-JUL-2001.  
XX 22-SEP-2000; 2000GB-0023315.  
XX 23-SEP-1999; 99US-0155493.  
PR 23-SEP-1999; 99US-0404133.  
PR 23-SEP-1999; 99US-0404133.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX (PHAA ) PHARMACIA & UPJOHN CO.  
XX Bienkowski MJ, Gurney M;  
XX WPI; 2001-444208/48.  
XX N-PSDB; AAD17865.  
XX Polypeptide comprising fragments of human aspartyl protease with  
XX amyloid precursor protein processing activity and alpha-secretase  
XX activity, for identifying modulators useful in treating Alzheimer's  
XX disease -  
XX Example 2; Fig 2; 187pp; English.  
XX The patent discloses human aspartyl protease 1 (hu-Asp1) or modified  
XX Asp1 proteins which lack transmembrane domain or amino terminal  
XX domain or cytoplasmic domain and retains alpha-secretase activity  
XX and amyloid protein precursor (APP) processing activity. The proteins  
XX of the invention are useful for assaying hu-Asp1 alpha-secretase  
XX activity, which in turn is useful for identifying modulators of  
XX hu-Asp1 alpha-secretase activity, where modulators that increase  
XX hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's  
XX disease (AD) which causes progressive dementia with consequent  
XX formation of amyloid plaques, neurofibrillary tangles, gliosis and  
XX neuronal loss. Hu-Asp1 protease substrate is useful for assaying  
XX hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with  
XX the substrate under acidic conditions and determining the level of  
XX hu-Asp1 proteolytic activity. The present sequence is long form of  
XX human Asp2 protein, designated as Asp2(a). Asp2 gene is localised  
XX on chromosome 11q23.3-24.1.  
XX Sequence 501 AA;

Query Match 99.88; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.88; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ETDEPEEPGRGSGFVEMVDNLRGKSGQGYVEMTVGSPPTNLIVDTGSSNFAVGAAP 60  
Db 46 etdeepeepgrgsgfvmvndnlgksgggvymtvgsppqtnilvdtgssnfavgaap 105  
QY 61 HPFLHRYRQRLSTYRDLRKGVVYPTQKWEGETDLVSIPIHGPNTVVRANIAITE 120  
Db 106 hpflhryyrglsltyrdlrgvvyptqgkwegeltvlsvipghpntvvranaiaite 165  
QY 121 SDKEFFINGNWEGLGLAYAEIARPDLSLPPFDLSLVKQTHVPLFSLQLCGAGFPINQS 180  
Db 166 sdekffingsnweglglayaeiarpdslslepfdslvkvqthvpnlflhlcgagfpinqs 225  
QY 181 EVLASVGSMIIGIDHSLXTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDK 240  
Db 226 evlasvggsmiigidhsltygslwypirrewyevliivrveingqdlkmdckeynydk 285  
QY 241 SIYDSGTNLRPKKVEAAVSKIAASSTEKPPDGLWGLQVLCWQAGTTPNIFPVIS 300  
Db 286 siydsrgtnlrpkkvteaaavksikaasstekpdpdglwglqvlcwqagtpnifpvvis 345  
QY 301 LYLMGEVNTGSRFTIIPQYLRPVEDVATSDDCYKFAISQSSTGVMGAVIMEGYVYV 360  
Db 346 lylmgevtngsfrtliipqyirpvedvatsdddcykfaissqstgtvmgavimegyfvv 405

QY 361 FDRARKRIGFAVSACHVHDEFTAAVEGPFVTLQMEDCGYNIPQDDESTLMTIAYVMAAI 420  
Db 406 fdrarkrigrifavsvachvhdeftaavegpfvtldmedcgyntpqtdestlmtiayvmaai 465  
QY 421 CALEMLPLCLMVCOWRCRLCRLQHQHDDFADDISLLK 456  
Db 466 calemlplclmvcowrcrlcrlrqhddiaddisllk 501  
RESULT 10  
AAE06859  
ID AAE06859 standard; Protein; 501 AA.  
XX AAE06859;  
XX 23-OCT-2001 (first entry)  
XX Human aspartyl protease 2a (Hu-Asp2a) protein.  
XX Human; aspartyl protease 2a; Asp 2a; beta-amyloid precursor protein; APP;  
XX beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;  
XX neurofibrillary tangle; neuronal loss; amyloid-beta peptide; nootropic;  
XX neuroprotective; antisense therapy; gene therapy;  
XX chromosome 11q23.3-24.1.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Peptide 1..21 /label= Signal\_peptide  
XX Protein 22..501 /label= Mature human aspartyl protease 2a (Hu-Asp2a)"  
XX Region 420..454 /note= "Alpha helical spacer region"  
XX Domain 455..477 /label= Transmembrane\_domain  
XX Domain 478..501 /label= Cytoplasmic\_domain  
XX WO200150829-A2.  
XX 19-JUL-2001.  
XX 09-MAY-2001; 2001WO-IB00799.  
XX 09-MAY-2001; 2001WO-IB00799.  
XX (BIEN/) BIENKOWSKI M J.  
XX (GURN/) GURNEY M E.  
XX (HEIN/) HEINRIKSON R L.  
XX (PARO/) PARODI L A.  
XX (YANR/) YAN R.  
XX Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX WPI; 2001-483072/52.  
XX N-PSDB; AAD13021.  
XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
XX protease 2, lacking Asp2 transmembrane domain and retaining beta  
XX secretase activity of Asp2 useful for identifying inhibitors of Asp2  
XX activity -  
XX Claim 49; Fig 2; 185pp; English.  
XX The invention relates to human aspartyl proteases (Hu-Asp), beta-amyloid  
XX precursor protein (APP) isoforms and their corresponding DNA molecules.  
XX Human aspartyl proteases can act as beta-secretase proteases useful for  
XX treating Alzheimer's disease. APP isoforms are useful for identifying  
XX modulators of amyloid-beta peptide production, for use in designing  
XX therapeutics for the treatment and prevention of Alzheimer's disease,  
XX dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis

CC and neuronal loss. App isoforms are also used in methods for identifying  
CC inhibitors and modulators of human Asp2 activity. The invention relates  
CC to a method for identifying agents that modulate the activity of human  
CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
CC as a means to screen in cellular assays for the inhibitors of beta- and  
CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
CC polymerase chain reactions (PCR). The probes are useful for detecting  
CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
CC blots. The present sequence is human aspartyl protease 2 (Hu-Asp2), a  
CC 'long' form designated as (Hu-Asp2a). Hu-Asp 2 gene is localised on  
CC chromosome 11q23.3-24.1.  
XX  
XX  
SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGYVEMTVGSPPTNLILVDTGSSNFAVGAAP 60  
Db 46 etdeepeepgrgsfvmvndnlrgsggyvemtvgspptnlilvdtgssnfavgaap 105

QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWKEGELGDLVSIHPGPNVTVRANIAATE 120  
Db 106 hpflhryyqrlsstyrdlrgvvyptqgkwegeldlvsiphgpnvtvranaiaate 165

QY 121 SDKFFINGNWEGILGLAYARIAPDPSLEFFDLSLVKQTHVPNLFSLQLCGAGFFLNQS 180  
Db 166 sdkffingsnwegilglayaeiarpdpsleffdsllvkqthvpnlflslhcgagfplngs 225

QY 181 EVLASVGSMTIGGIDHSLYTSLWYTPIRREWYVEVLIIVRVEINGDLMKDCKEYNVYDK 240  
Db 226 evlasvgsmtiggidhslytsglwytpirrewyvevliivrveingdldmkdkeynydk 285

QY 241 SIVDSGTTNLRPLPKKVFEEAAYKSIKAASSTTEKFPDGLWGLQVLCWAGTTPWNIFFVIS 300  
Db 286 sivdsgttnlrplpkkvfeavvksikaasstekfpdgfwlgeqlvcwaggttwnifpvis 345

QY 301 LYLMEVTNQSFRITILPQOYLREVEDVATSQDCYKFAISOSGTVMGAVIMEGFYVV 360  
Db 346 lylmevtnqsfrtilpqoylrvedvatSQDCYKFAISOSGTVMGAVIMEGFYVV 405

QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvltlmedcgynipctdestlmtiayvmaai 465

QY 421 CALFMLPCLMVCQWRCRLRQHQHDFADDSILK 456  
Db 466 calfmlpclmvcwrcrlrqhqhdfaddsilK 501

RESULT 11  
AAU06603  
ID AAU06603 standard; Protein; 501 AA.  
XX  
AC AAU06603;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human Aspartyl protease 2(a), Asp2(a).  
XX  
KW Human; Aspartyl protease; Asp2(a); beta-secretase; neurotropic;  
KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
KW amyloid-beta; Abeta.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT Peptide /label= Signal peptide  
FT Peptide 22..45  
FT Peptide /label= Pre\_pro\_peptide

Peptide 46..57  
FT /label= Pro\_peptide  
FT Protein 57..501  
FT /label= Mature\_Asp2(a)  
FT Region 420..454  
FT Domian 455..477  
FT /label= Transmembrane\_domain  
FT Domian 478..501  
FT /label= Cytoplasmic\_domain  
XX  
PN WO200149098-A2.  
XX  
XX 12-JUL-2001.  
XX  
PD 09-MAY-2001; 2001WO-IB00798.  
PF  
XX 09-MAY-2001; 2001WO-IB00798.  
PR  
XX (BIEN/) BIENKOWSKI M J.  
PA (GURN/) GURNEY M E.  
PA (HEIN/) HEINRIKSON R L.  
PA (PARO/) PARODI L A.  
PA (YANR/) YAN R.  
XX  
PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX  
XX WPI: 2001-502549/55.  
DR N-PSDB; AAS11517.  
XX  
XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity  
XX  
PS Claim 49; Fig 2; 185pp; English.  
XX  
XX The invention relates to a purified polypeptide comprising a fragment of  
CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2  
CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
CC the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. The invention also details polynucleotides for the Asp  
CC proteins and vectors expressing them, and a polypeptide (isoform of  
CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
CC APP or its fragment containing an APP cleavage site recognizable by a  
CC mammalian beta-secretase, and further comprising two lysine residues at  
CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
CC APP fragment. Also included in the invention are methods of identifying  
CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
CC useful for treating Alzheimer's disease. APP is useful in methods for  
CC identifying inhibitors or modulators of human Asp2 activity and  
CC therapeutics for the treatment or prevention of Alzheimer's disease.  
CC APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which  
CC is associated with increased levels of Abeta processing is useful in  
CC assays relating the Alzheimer's research. The expression vector is useful  
CC for recombinantly expressing APP. Nucleic acids that hybridise to  
CC Asp oligonucleotides are useful as probes or primers. The probes are  
CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence is human Asp2(a).  
XX  
SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGYVEMTVGSPPTNLILVDTGSSNFAVGAAP 60  
Db 46 etdeepeepgrgsfvmvndnlrgsggyvemtvgspptnlilvdtgssnfavgaap 105

QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWKEGELGDLVSIHPGPNVTVRANIAATE 120



Db 106 hpfihryyqrqlsystyrdlrgkgyvpytqgkwegelgtdivsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGTLGLAYAEIARPDPSLEPPFDLSLVKQTHVPLNLSQLCGAGFPLNQS 180  
Db 166 sckffingsnwegilglayaeiarppdslepfdslvkqthvpnlfslhicgagfplnqs 225  
QY 181 EVLASVGSMIIIGIDHSLTGSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNVDK 240  
Db 226 evlasvggsmiigidhsltytgslywtpirrewyevliivrvveingqdlkmdckeynydk 285  
QY 241 SYVDSGTTNLRPKKVFEEAAVKSIKAASSTEKFPDGFGLVGEQLVCWQAGTTPWNIFPVIS 300  
Db 286 sivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqaggttwnifpvvis 345  
QY 301 LYLWGEVTNQSFRITILPQOYLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV 360  
Db 346 lylwgevtnqsfrilitlpqgylrpvedvatsqddcykfaissqstgtvmgavimegfyyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgyinipqtdestlmtiayvmaai 465  
QY 421 CALFMLPLCLMVCQWRCLRLRQOHDDFADDISLLK 456  
Db 466 calfmplclmvcqwrclrlrqghddfaddisilk 501

RESULT 12

AAU07202

ID AAU07202 standard; Protein: 501 AA.

XX AC AAU07202;

XX DT 24-OCT-2001 (first entry)

XX DE Human aspartyl protease 2a (Asp-2a).

XX KW Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;  
KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;  
KW beta-secretase; Alzheimer's disease.  
XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT Peptide 1..21  
FT /note= "Signal peptide"  
FT Misc\_feature 22..45  
FT /note= "Pre-propeptide"  
FT Misc\_feature 46..57  
FT /note= "Propeptide"  
FT Protein 58..501  
FT Region 420..454  
FT /note= "Mature Aspartyl protease-2a"  
FT Domain 455-477  
FT /note= "Alpha helical spacer region"  
FT Domain 478..501  
FT /note= "Transmembrane domain"  
FT /note= "Cytoplasmic domain"

XX WO200149097-A2.

XX 12-JUL-2001.

XX PF 09-MAY-2001; 2001WO-IB00797.

XX PR 09-MAY-2001; 2001WO-IB00797.

XX PA (BIEN/) BIENKOWSKI M J.

XX PA (GURN/) GURNEY M E.

XX PA (HEIN/) HEINRIKSON R L.

XX PA (PARO/) PARODI L A.

XX PA (YANR/) YAN R.

XX PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX WPI: 2001-502548/55.  
XX N-PSDB; AAS11702.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity

XX Claim 49; Fig 2: 185pp: English.

XX The invention relates to a novel purified polypeptide comprising a  
CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
CC and the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. Also included is an isoform of amyloid protein precursor (APP)  
CC comprising the amino acid sequence of a APP or its fragment containing  
CC an APP cleavage site recognisable by a mammalian beta-secretase, and  
CC further comprising two lysine residues at the carboxyl terminus of the  
CC amino acid sequence of the mammalian APP or APP fragment. The  
CC polypeptides are used for assaying for modulators of beta-secretase  
CC activity; identifying agents that inhibit the APP processing activity  
CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that  
CC modulate the activity of Asp2; and for reducing cellular production of  
CC amyloid beta (Abeta) from APP. Agents identified by the above methods  
CC are useful for treating Alzheimer's disease; and for identifying  
CC modulators of amyloid-beta (Abeta) peptide production, for use in  
CC designing therapeutics for the treatment or prevention of Alzheimer's  
CC disease. Probes and primers derived from Asp nucleic acid sequences  
CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence represents the  
CC amino acid sequence of human Asp-2a used in the methods of the invention.

XX SQ Sequence 501 AA;

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEEPEEPGRGRGSFVEMVDNLRGKSGQYVEMTVGSPPTLNILVDTGSSNFAVGAAP 60  
Db 46 etdeepeepegrgrgsfvmvndnlrgksggyvemtvgspptlnilvdtgssnfavgaap 105  
QY 61 HPFLHRYYQRLSSTYRDLRKGVVYVYTGQKWEGETGLDLSVPHGPNVTVRANIAITE 120  
Db 106 hpflhryyqrqlsystyrdlrgkgyvpytqgkwegelgtdivsiphgpnvtvranaiaite 165  
QY 121 SDKFFINGSNWEGTLGLAYAEIARPDPSLEPPFDLSLVKQTHVPLNLSQLCGAGFPLNQS 180  
Db 166 sckffingsnwegilglayaeiarppdslepfdslvkqthvpnlfslhicgagfplnqs 225  
QY 181 EVLASVGSMIIIGIDHSLTGSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNVDK 240  
Db 226 evlasvggsmiigidhsltytgslywtpirrewyevliivrvveingqdlkmdckeynydk 285  
QY 241 SYVDSGTTNLRPKKVFEEAAVKSIKAASSTEKFPDGFGLVGEQLVCWQAGTTPWNIFPVIS 300  
Db 286 sivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqaggttwnifpvvis 345  
QY 301 LYLWGEVTNQSFRITILPQOYLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGFYVV 360  
Db 346 lylwgevtnqsfrilitlpqgylrpvedvatsqddcykfaissqstgtvmgavimegfyyv 405  
QY 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420  
Db 406 fdrarkrigfavsachvhdefrtaavegpfvtldmedcgyinipqtdestlmtiayvmaai 465  
QY 421 CALFMLPLCLMVCQWRCLRLRQOHDDFADDISLLK 456  
Db 466 calfmplclmvcqwrclrlrqghddfaddisilk 501

```
RESULT 13
AAE02581
ID AAE02581 standard; Protein; 501 AA.
XX
AC AAE02581;
XX
DT 10-AUG-2001 (first entry)
XX
DE Human aspartyl protease 2a (Asp 2a).
XX
KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;
KW Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp 2a;
KW beta-secretase; chromosome 11q23.3-24.1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Signal_peptide
FT Peptide 22..45
FT /label= Asp_2a_prepropeptide
FT Peptide 46..57
FT /label= Asp_2a_propeptide
FT Protein 58..501
FT /label= Mature_human_Asp_2a_protein
FT Active-site 93..95
FT /label= Active_site_1
FT Active-site 289..291
FT /label= Active_site_2
FT Region 420..454
FT /label= Alpha_helical_spacer
FT Domain 455..477
FT /label= Transmembrane_domain
FT Domain 478..501
FT /label= Cytoplasmic_domain
FT Region 486..501
FT /note= "Peptide #2"
XX
PN WC200123533-A2.
XX
XX 05-APR-2001.
XX
XX 22-SEP-2000; 2000WO-US26080.
XX
XX 23-SEP-1999; 99US-0155493.
XX 23-SEP-1999; 99WO-US20881.
XX 13-OCT-1999; 99US-0416901.
XX 06-DEC-1999; 99US-0169232.
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX Gurney M, Bienkowski MJ;
XX
XX WPI; 2001-290516/30.
XX N-PSDB; AAD06739.
XX
XX Enzymes that cleave the alpha-secretase site of the amyloid precursor
XX protein, useful for the treatment of Alzheimer's disease -
XX
XX Example 2; Fig 2; 189pp; English.
XX
XX The present invention relates to enzymes for cleaving the alpha-
XX secretase site of the amyloid precursor protein (APP) and methods of
XX identifying those enzymes. The methods may be used to identify enzymes
XX that may be used to cleave the alpha-secretase cleavage site of the APP
XX protein. The enzymes may be used to treat or modulate the progress of
XX Alzheimer's disease. The present sequence is human aspartyl protease 2a
XX (Asp 2a). Asp 2a has beta-secretase protease activity. Asp2 gene
XX is located on chromosome 11q23.3-24.1.
XX
XX Sequence 501 AA;
```

```
Query Match 99.8%; Score 2414; DB 22; Length 501;
Best Local Similarity 99.8%; Pred. No. 1.5e-241;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ETDDEPEPEGRGRGSFVEMVDNLRGKSGQGYVEMTGVSPQPTNLILVDTGSSNFANGAAP 60
Db 46 etdeepeepgrgrgsfvmvndlrgksqgyvemtgvspqptnlilvdtgssnfavgaap 105
QY 61 HPFLHRYQRQLSSTYRDLRKGVVVPYTOCKWEGELGTLVSIPIHGPNVTVRANIAITE 120
Db 106 hpfhlrhyqrqlsstyrdlrkgyvvyptgkwegeigtldlvsiphgpnvtvranaiaite 165
QY 121 SDRFFFNNGSWEIGLGLAYAEIARPDSDLPEFFDLSLVKQTHVPNLFSLQLCGAGFPLNQS 180
Db 166 sdhffingsnwegilglayaeiarpddslepfdslvkqthvpnlfslhlcgagfplnqs 225
QY 181 EVLASVGSMSIIGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNKYDK 240
Db 226 evlasvggsmilggidhsltytgslywtpirrewyveilvrveingqdlkmdckeynydk 285
QY 241 SIVDSGTTNLRPKKVFEEAAVKSIAASSTKPPDGFMLGEOLVVCWQAGTTPWNIFPVIS 300
Db 286 sivdsgettlnlrpkkvfeaaavksikaasstekfpdgtfwlgeqlvcwaggttpwnifvis 345
QY 301 LYLMGEVTNQSFRITILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 346 lylmgevtngsfrtilpqqylrpvedvatsqdqdcykfaissgstgtvmgavimegfyyv 405
QY 361 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDSTLMTIAYVMAAI 420
Db 406 fdrarkrigfavsachvhdefrtaaavegpfvtldmedcgyinipqtdstlmtiayvmaai 465
QY 421 CALFMPLCLMVCQWRCLRCLROQHDDFADDISLLK 456
Db 466 calfmplclmvcwqwrclrlrqhddfaddisllk 501
RESULT 14
AAW59807
ID AAW59807 standard; Protein; 501 AA.
XX
AC AAW59807;
XX
DT 26-OCT-1998 (first entry)
XX
DE Amino acid sequence of human ASP2 (aspartic protease 2).
XX
KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
KW prohormone processing.
XX
XX Homo sapiens.
XX
XX PN EP855444-A2.
XX
XX 29-JUL-1998.
XX
XX 27-JAN-1998; 98EP-0300573.
XX
XX 28-JAN-1997; 97GB-0001684.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Chapman CG, Murphy K, Powell DJ, Smith TS;
XX
XX WPI; 1998-389809/34.
XX DR N-PSDB; AAV41696.
XX
XX New nucleic acid encoding human aspartic protease 2 - used to treat,
XX prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
XX PT
```



Search completed: August 7, 2002, 09:16:14  
Job time: 159 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 32.31 Seconds  
(without alignments)  
1356.135 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEEPEEPRGRGSFVEMVD.....CLRLRQHQHDDFADDSILK 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2419	100.0	501	2 A59090	aspartic proteinase
2	308	12.7	384	2 JC7574	pepsinogen A - Afr
3	308	12.7	385	2 JC7575	pepsinogen A - bul
4	306	12.6	387	2 B38302	pepsin (EC 3.4.23)
5	303	12.5	383	2 JC7573	pepsinogen C - Afr
6	302.5	12.5	388	1 S19082	pepsin A (EC 3.4.2
7	301	12.4	382	1 PECH	pepsin A (EC 3.4.2
8	299.5	12.4	396	2 A34401	cathepsin E (EC 3
9	298.5	12.3	383	2 A41443	pepsin (EC 3.4.23)
10	298.5	12.3	384	2 A39314	gastricsin (EC 3.4
11	296	12.2	387	2 C38302	pepsin (EC 3.4.23)
12	296	12.2	391	2 A43356	cathepsin E (EC 3
13	295.5	12.2	412	1 KHHUD	cathepsin D (EC 3
14	295	12.2	387	2 D38302	pepsin (EC 3.4.23)
15	289.5	12.0	388	1 S19684	pepsin (EC 3.4.23)
16	287.5	11.9	444	2 T24204	hypothetical prote
17	287	11.9	407	1 KHRTD	cathepsin D (EC 3
18	286	11.8	387	2 E38302	pepsin (EC 3.4.23)
19	285.5	11.8	398	2 S66465	cathepsin E (EC 3
20	285	11.8	398	2 S11185	cathepsin D (EC 3
21	283.5	11.7	388	1 PEMQAR	pepsin A (EC 3.4
22	282.5	11.7	388	1 PEHU	pepsin A (EC 3.4
23	281.5	11.6	381	1 CMSEB	chymosin (EC 3.4
24	281.5	11.6	388	1 PEMQAJ	pepsin A (EC 3.4
25	281.5	11.6	388	2 A30142	pepsin A (EC 3.4
26	279.5	11.6	388	2 B30142	pepsin A (EC 3.4
27	279.5	11.6	410	1 KHM5D	cathepsin D (EC 3
28	278.5	11.5	386	1 PEFG	pepsin A (EC 3.4
29	278	11.5	387	2 JC7245	pepsinogen A - com

ALIGNMENTS

```
RESULT 1
A59090
aspartic proteinase (EC 3.4.23.-) BACE precursor - human
N;Alternate names: beta-secretase, beta-site APP cleaving enzyme
C:Species: Homo sapiens (man)
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-May-2000
C:Accession: A59090
R;Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplo
M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro
Science 286, 735-741, 1999
A:title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran
A:Reference number: A59090; MUID:20002972
A>Note: submitted to GenBank, September 1999
A:Accession: A59090
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-501 <VAS>
A:CROSS-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539
C:Genetics:
C:Gene: BACE
C:Superfamily: beta-secretase
C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-45/Domain: propeptide #status predicted <PRO>
F:46-501/Product: acid proteinase BACE #status predicted <MAT>
F:461-477/Domain: transmembrane #status predicted <TRN>
F:93,289/Active site: Asp #status predicted
F:153,172,223,354/Binding site: carboxylate (Asn) (covalent) #status predicted
F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2419; DB 2; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.9e-192;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETDEEPEEPRGRGSFVEMVDNLGRKSGQYGVVPTQGWEGELGTDLVSTPHGPNVTVRANIAAITE 120
Db 46 ETDEEPEEPRGRGSFVEMVDNLGRKSGQYGVVPTQGWEGELGTDLVSTPHGPNVTVRANIAAITE 105
QY 61 HPFLHRYQRQLSSTYRDLRKGVVPTQGWEGELGTDLVSTPHGPNVTVRANIAAITE 120
Db 106 HPFLHRYQRQLSSTYRDLRKGVVPTQGWEGELGTDLVSTPHGPNVTVRANIAAITE 165
QY 121 SKFFINGNWEIGILGLAYAEIARPDSDLEPFDSLVKQTHVPNLFSLQCGAGPPLNGS 180
Db 166 SKFFINGNWEIGILGLAYAEIARPDSDLEPFDSLVKQTHVPNLFSLQCGAGPPLNGS 225
QY 181 EYLVASVGGSMIIGGDHSLYTGSLWYTPIRREMYVEVIVRVEINGQDLKMDCKEYNDK 240
Db 226 EYLVASVGGSMIIGGDHSLYTGSLWYTPIRREMYVEVIVRVEINGQDLKMDCKEYNDK 285
QY 241 SIIVDSGTTNLRPKKVFEEAAVKSIAAASSTKFPDGFGLGEQLVCWQAGTTPWNIFPVIS 300
```

```

|||||
Db 286 SIYDSGTTNLRPKKFEAAVSIKAASTEKEFPDGFGEQLVQCQAGTTPNIFPVIS 345
QY 301 LYLMEVTNQSFRITLPPQYLPRVEDVATSQDDCYKFAISQSTGTVMGAVIMEGFYV 360
|||||
Db 346 LYLMEVTNQSFRITLPPQYLPRVEDVATSQDDCYKFAISQSTGTVMGAVIMEGFYV 405
QY 361 FDRARKRIGFAVACHVHDFRFAAEGPVTILDMEDCGYNIPQTDSTLMTIAYVMAAI 420
|||||
Db 406 FDRARKRIGFAVACHVHDFRFAAEGPVTILDMEDCGYNIPQTDSTLMTIAYVMAAI 465
QY 421 CALFMLPLCLMVCQWRCLRLRQOHDDFADDISLLK 456
Db 466 CALFMLPLCLMVCQWRCLRLRQOHDDFADDISLLK 501

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7574; PC7119
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Cross-references: DDBJ:AB045380
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Molecule type: protein
A:Residues: 16-35;57-76 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: PgA
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 12.7%; Score 308; DB 2; Length 384;
Best Local Similarity 27.5%; Pred. No. 1.3e-17;
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;
QY 30 YVEMTVGSPPTLNILVDGTSSNFVAGPFL-----HRYQQLSSTYRDLRKG 82
||| ::::|::| ::|::|::| ::|::|::| ::|::|::| ::|::|::|
Db 72 YGTISIGTPPQEFVIFDTGSANLW---PSVYCSCQACSNHNRFPQOSTFQATNP 128
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 83 VVVPYTGKGWEGELGTDLSIPHPNVTVRANIAAITESDK--FFINGSNWEGILGLAYAE 141
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 129 VSIQYGTGSMGFLGYDTLQV---GNIQISNQMFGLSEPEGSFLYISPFDDGILGLAFPS 185
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 142 TARPDDSLPEFDSLVKQTHVP--NLFSLOLCGAGFPLNQSEVLASVGSMTIIGDHSLY 200
||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 186 IA--SSQATPVFDNMWSOGLIPQNLFSVYSSDG-----OTGSYVLFGGVDNSYY 233
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 201 TGSILWYTPIRREWEYEVIIVRVEINGQDL--KMDCKEYNDKSIDVSGTTNLRPLPKVFE 258
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 234 SGLSNWPLTAETWQITLDSVINGQVACSQSC-----QAIVDGTSLMTGPSTPI- 286
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 259 AAVSIKAASTEKEFPDGFGEQLV-CWQAGTPWNIPFVISLYLMEVNTQSFRTIL 317
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 287 ANIQNYIGASQDSN-----GQVYINCNISNMPTIVF-----TIN 321
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 318 PQQY-LRPVEDVATSQDDCYK--FAISQSSTGT-----VMGAVIMEGFYVDFRARRKRGFA 371
||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 322 GVQYPLPSAYVRQNOQCGSSGFQAMNLPNTSGDLWILGDVFIROYTFVFDNRANNVVAIA 381
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
JC7575
pepsinogen A - bullfrog

```

```

C:Species: Rana catesbeiana (bullfrog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7575
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Cross-references: DDBJ:AB045376
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: PgA
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 12.7%; Score 308; DB 2; Length 385;
Best Local Similarity 27.8%; Pred. No. 1.3e-17;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;
QY 30 YVEMTVGSPPTLNILVDGTSSNFVAGPFL-----AAPHFLHRYQQLSSTYRDLRKGVV 85
||| ::::|::| ::|::|::| ::|::|::| ::|::|::| ::|::|::|
Db 73 YFTISIGTPPQSTFVIFDTGSANLWPSVYCSPACTNHHMFNQOSTFQATNPVSI 132
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 86 PYTGKGWEGELGTDLSIPHPNVTVRANIAAITESDK--FFINGSNWEGILGLAYAEIAR 144
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 133 QYGTGSMGFLGYDTVQV---GNIQITNQIFGLSQSEPGSFLYISPFDDGILGLAFPSLA- 188
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 145 PDDSLPEFDSLVKQTHVP--NLFSLOLCGAGFPLNQSEVLASVGSMTIIGDHSLYTGS 203
||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 189 -SSQATPVFDNMWNOGLIPQDLFSVYSSQG-----GSFVLFGGVDTSYTGN 237
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 204 LWTPTIRREWEYEVIIVRVEINGQDLKM--DCKEYNDKSIDVSGTTNLRPLPKVFEAAV 261
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 238 LNWVPLTAETWQITLDSVINGQVACSQSC-----SAIVDTGSLLAGP----STPI 287
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 262 KSIKAASTEKEFPDGFGEQLVQCQAGTPWNIPFVISLYLMEVNTQSFRTILPQQY 321
||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 288 ANIQYIGANQDSNGQYV---INCNNISNMPTVVF-----TINGVQY 326
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 322 LRPVED-VATSQDDC---YKFAISQSSTGT--VMGAVIMEGFYVDFRARRKRGFA 371
||| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 327 PLPASAYVROSQOSTSGFQAMNLPNTSGDLWILGDVFIREYTVFVFDNRANNVAMA 382
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 4
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.6%; Score 306; DB 2; Length 387;
Best Local Similarity 27.1%; Pred. No. 2e-17;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;
QY 30 YVEMTVGSPPTLNILVDGTSSNFVAGPFL-----AAPHFLHRYQQLSSTYRDLRKGVV 85
||| ::::|::| ::|::|::| ::|::|::| ::|::|::| ::|::|::|

```

Dd 75 YFGTISIGPPQEVFTDGTSSNLWVPSTYSSSLACFLHKKRFPDSDSTFQATSETLSI 134  
QY 86 PYTOQKWEGLGELTDLVSIPIHGPNTVRANIAAITESD---KEFINGSNWEGLGLAYAEI 142  
Dd 135 TYTGSMTGILGYDIVKV---GNIEDTNGIFGLSKTEPITFLV--APFDGILGLAYPSI 189  
QY 143 ARPDSLEFPFSLVKQTHV--PNLFSLQLCGAGFPLNQSEVLASVGGSMIIGDHSLYT 201  
Dd 190 SASDAT--PVFDNMWNEGLVSEDLFSVYLSNG-----EKGSMVMFGGIDSSYYT 237  
QY 202 GSLWTPPIRREWYEVIIIVRVEINGQDLKM--DCKEYNYDKSIIVDSGTTNLRPLPKVFEA 259  
Dd 238 GSLNWPVSHGTYWQITMDISITINGIACADSC-----QAVDGTGTLGAGPTSATSK 291  
QY 260 AVKSIKAASSTKFPDGFGLGQLV-CWQAGTTPNIFPVVISLYLMGEVNTQSFRTITLP 318  
Dd 292 IQSYIGASKNL-----LGENIISCSAIDSLDIVF-----TINN 325  
QY 319 QOYLRPVED-VATSDDC---YKFAISQSSTGT--VMGAVIMEGFYVVDRAKRGIFAV 372  
Dd 326 VOYPLPASAYILKEDDCCLSGFDGMNLDTSYGELWILGDFVIRQYFTVFDRAANNQVGLAA 385  
QY 373 SA 374  
Dd 386 AA 387

RESULT 5  
JC7573

pepsinogen C - African clawed frog  
N:Alterate names: progastricsin  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7573; PC7118  
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.  
J. Biochem. 129, 147-153, 2001  
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens  
A:Reference number: JC7573; MUID:21064922; PMID:11134969  
A:Contents: Stomach  
A:Accession: JC7573  
A:Molecule type: mRNA  
A:Residues: 1-383 <IKU>  
A:Cross-references: DBJ:AB045379  
A:Accession: PC7118  
A:Molecule type: protein  
A:Residues: 17-68 <IK2>  
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a  
C:Genetics:  
A:Gene: Pgc  
C:Superfamily: pepsin  
C:Keywords: stomach; zymogen

Query Match 12.5%; Score 303; DB 2; Length 383;  
Best Local Similarity 27.4%; Pred. No. 3.4e-17;  
Matches 104; Conservative 57; Mismatches 122; Indels 96; Gaps 17;

QY 30 YVEMTVGSPQTLNILDVDTGSSNFAVGA-----APHPFLHRYQRLSSTYDLRK 81  
Dd 67 YGEISICTPQNFVLVFDGTSSNLWASTYCSQACVNHPL----FNPSQSTYSSNQ 122  
QY 82 GYVPYTOGKWEGLGELTDLVSIPIHGPNTVRANIAAITESDKFFINGSN-----WEGILG 136  
Dd 123 QFSLOYGTGSLTGILGYDTVTI---QNVAISQOERGLSETEP---GTFNVYAQFDGILG 175  
QY 137 LAYAEIARPDSDSLEFPFSLVKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGID 196  
Dd 176 LAYPSIA--VGATVVMGMMQ-----NLLNQPI--FGYLSGQS--SONGGEVAFGGVD 225  
QY 197 HSLYTGSLWYTPIRREWYEVIIIVRVEINGQD---LKMCKEYNYDKSIIVDSGTTNLRPL 253  
Dd 226 QNYITGYIYWPVTSETWQIGQFSGINGQATGWCSGC-----QAIVDTGTSLLTAP 279

QY 254 KVFEEAAVKSIAASSTKFPDGFGLGQLVCMQAGTTPWNI--RPVISLYLMG----- 305  
Dd 280 QSVFSSLIQISGAQDON-----GQVVSCH-----NIQNLPITISFTISGVSPPLP 325  
QY 306 ---EVTNQS-----FRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 355  
Dd 326 PSAYVLQSSGYCTIGIMPTYLPSONGQPL-----WILGDVFLR 364  
QY 356 GFYVVDRAKRGIFAVSA 374  
Dd 365 EYISYIDLGNNOVGFEATAA 383

## RESULT 6

S19682  
pepsin A (EC 3.4.23.1) 4 precursor - Japanese macaque  
N:Alterate names: pepsinogen A isozyme 4  
C:Species: Macaca fuscata (Japanese macaque)  
C:Date: 22-Nov-1993 #sequence\_revision 19-Oct-1995 #text\_change 18-Jun-1999  
C:Accession: S19682; S16065  
R:Kageyama, T.; Tanabe, K.; Koiwai, O.  
Eur. J. Biochem. 202, 205-215, 1991  
A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and st  
A:Reference number: S19681; MUID:92037645  
A:Accession: S19682  
A:Molecule type: mRNA  
A:Residues: 1-388 <KAG>  
A:Cross-references: EMBL:X59753; NID:g38070; PIDN:CAA42425.1; PID:g38071  
A:Note: parts of sequence, including amino ends of pepsinogen and activation intermed  
C:Comment: This is a minor component of pepsin at all post-partum stages.  
C:Comment: Although two-step activation is observed, activation is predominantly a o  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:16-388/Product: pepsinogen A #status experimental <PPT>  
F:16-62/Domain: activation peptide #status experimental <APT>  
F:63-388/Product: pepsin A #status experimental <BNZ>  
F:38-39/Cleavage site: Leu-Lys (pepsin) #status experimental  
F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental  
F:94-277/Active site: Asp #status predicted  
F:107-112,268-272,311-344/Disulfide bonds: #status predicted  
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 12.5%; Score 302.5; DB 1; Length 388;  
Best Local Similarity 27.6%; Pred. No. 3.8e-17;  
Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;

QY 3 DEEPEEPGRGSEVEMVDNLRGSGQGYVEMTVGSPQTLNILDVDTGSSNFAVGAAPHP 62  
Dd 64 DEQPLE----NYLDV-----EYFGTIGTGPAGNFTVVDTCSSNLWV---PSV 105  
QY 63 FL-----HRYQRLSSTYDLRKGVVVPYTOGKWEGLGELTDLVSIPIHGPNTVRANI 115  
Dd 106 YCYSLACMDHNLFPQDSSTYRATSKTVSYTGCSMTGILGYDTVKV---GGISDTNOI 162  
QY 116 RAITESDK--FFINGSNWEGLGLAYAEIARPDSDSLEFPFSLVKQTHV--PNLFSLQLCGA 173  
Dd 163 FGLSTEPGFLLYFAPFDGILGLAYPSIS--SSGATPVFDNINQRLVSDQLFSVYLSAD 220  
QY 174 GFPLNQSEVLASVGGSMIIGDHSLYTGSLWYTPPIRREWYEVIIIVRVEINGQDL--KM 231  
Dd 221 ----DQS-----GSVWIFGGIDSSYTGSLNWPVSVGYSWQISVDSITMNGKIACAK 270  
QY 232 DCKEYNYDKSIIVDSGTTNLRPLPKVFEAAVKSIAASSTKFPDGFGLGELV--CWQAGT 290  
Dd 271 GC-----QAIVDTGTSLLTGPTSPIANIQSDIGASENSD-----GEMVVSCTAIS 316  
QY 291 TPWNIFPVISLYLMGEVNTQSFRTILPQOY--LRPVEDVATSDDCYK-----FAISQSS 344  
Dd 317 LPDIVF-----TINGVQYPLPPPSAYILQSGSGCTSGFGQGMVPTBSG\_358  
QY 345 TGTVMGAVIMEGFYVVDRAKRGIFA 371





Db 170 TEPGQTEVD-AEFDGILGLGYPSLA--VGGVTPVFDNNMAQ-----NLVDLPMSFVMSNN 222  
QY 179 QSEVLASVGGSMIIIGGDHSLYTGSLAWTPIRREWYVEIIVRVEINGODLKMDCKEYNY 238  
Db 223 PE---GGAGSELIFGGDHSFGLSNWVPYKQAYQOIALDNIOVGG--TVMFCSE--G 275  
QY 239 DKSIVDSGTTNLRPKPKVFAAIVKSIKAASTKFPDGFGLGEOLVCWQAGTTPTWNIFFPV 298  
Db 276 CQAIIVDTGTSITGPSDKIKOLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322  
QY 299 ISLYLMGEVTVNQSPRIITLPOQYLRPVEDVATSDQCKYFAISOSSRG----- 346  
Db 323 VTFTING-----VPYTLSPAY--TLLDFVDGMQFC-----SSGFGQLDIHPGAP 366  
QY 347 -TVMGAVIMEGFYVVDRAKRKIGFA 371  
Db 367 LWILGDVFIQFYSVDFDGNRVGLA 392

## RESULT 9

A41443  
pepsin (EC 3.4.23.-) precursor, embryonic - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 21-Jul-2000  
C:Accession: A41443  
R:Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.  
J. Biochem. 103, 290-296, 1988  
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pep  
A:Reference number: A41443; MUID:88227903  
A:Accession: A41443  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-383 <HAY>  
A:Cross-references: GB:D00215; NID:g2760810; PIDN:BAA00153.1; PID:g222853  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 12.3%; Score 298.5; DB 2; Length 383;  
Best Local Similarity 25.2%; Pred. No. 8e-17;  
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY 30 YYVEMTVGSPPTNLINLVDTGSSNFAYGA----APHFLHRYQRLSSYRDLRKGVYV 85  
Db 76 YGTISIGTTPQDFTVVDFTGSSNLWVPYSCTSPACQSHQMFNPQSSTYKSTGQNLSI 135  
QY 86 PYTGKWEGLGTDLVSIHPGNVTVRANTAAITESDKFFINGSNWEGILGLAYAEIARP 145  
Db 136 HYTGDMEGTVGCDTVTVASLMDTNQLFGLST-SEPGQFFVY-VKFDGILGLGYPSLAA- 192  
QY 146 DLSLEPPFDSLVRKQTHV-PNLFLSLQLCGAGFPLNQSEVLASVGGSMIIIGDHSLYTGS 204  
Db 193 -DGITPVDNWNESLLEQNLSVYLS-----REPMGSMVYVGGIDESYFTGSI 240  
QY 205 WYPIREWYVEIIVRVEINGODL--KMDCKEYNDKSIIVDSGTTNLRPKPKVFEAAVK 262  
Db 241 NWIPVSQYQWQISMDSIIVKQEIACSSG-----QAIDTGTSLVAGPASDINDIQS 294  
QY 263 SIKAASSTEFPPDGFGLGEOLVCWQAGTTWNIFPVLSL-----YLMGEVTVNQSPRIILP 318  
Db 295 AVGANQNT-----YGEYSV-----NCSHILAMPDVPVFIIGI----- 326  
QY 319 QOYLRLPVEDVA---TSQDCKYFAISQSSTGTVMGAVIMEGFYVVDRAKRKIGFA 371  
Db 327 -QY--PVPALAYTEONGQGTCTMSSSQNSADLWILGDVFIQFYSVDFDGNRVGLA 380

## RESULT 10

A39314  
gastricsin (EC 3.4.23.3) precursor - bullfrog  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 19-Jun-1992 #sequence\_revision 19-Jun-1992 #text\_change 22-Jun-1999

C:Accession: A39314  
R:Yakabe, B.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kag  
J. Biol. Chem. 266, 22436-22443, 1991  
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and  
A:Reference number: A39314; MUID:92042186  
A:Accession: A39314  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-384 <YAK>  
A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 12.3%; Score 298.5; DB 2; Length 384;  
Best Local Similarity 25.5%; Pred. No. 8.1e-17;  
Matches 100; Conservative 61; Mismatches 142; Indels 89; Gaps 15;

QY 14 SFVEMVDNLKSGOGYVEMTVGSPPTNLINLVDTGSSNFAYGAAPHPFL-----HR 66  
Db 51 NFATAFEPLANYMDMSYGEISIGTTPQNFLVLFDTGSSNLWV---PSTYCQSQACTNHP 107  
QY 67 YYQRLSSTYRDLRKGVYVPTQGWEGELGTDLVSIHPGNVTVRANIA-----AI 118  
Db 108 QFNPSQSSSYSSNQOQFSLQYGTGSLTGILGYDTVQI-----QNIATISQOEFGLSV 158  
QY 119 TESDKFFINGSNWEGILGLAYAEIARPDSDLEPFDSLVKQTHVFN-LFSLQLCGAGFPL 177  
Db 159 TEPGTNFVY-AQFDGILGLAYPSTA--EGGATTVMGMIQQNLINQPLFAYLSGQQNSQ 215  
QY 178 NQSEVLASVGGSMIIIGDHSLYTGSMLYTPIRREWYVEIIVRVEINGQD---LKMDC 234  
Db 216 N-----GGEVAFGGVDQNYYSQIYWTPTVSETYQWIGIGQFSVNGQATWCSCQC- 266  
QY 235 EYNDKSIIVDSGTTNLRPKPKVFAAIVKSIKAASTKFPDGFGLGEOLV-CWQAGTTPW 293  
Db 267 ----QGIVDTGTSLTAPQSQVFSLSMGIAQDQN-----GOYAVSCSNIQSLPT 313  
QY 294 NIFPVI-----SLYLMGEVTVNQ---SPRITILPQOYLRPVEDVATSDQCKYFAISQ 342  
Db 314 ISFTISGVSFPLPSPASVYVQNSGYCTIGIMPTYLPSONGQPL----- 356  
QY 343 SSTGTVMGAVIMEGFYVVDRAKRKIGFAVSA 374  
Db 357 ---WILGDVFLRQYYSVVDLGNNGVGFAAAA 384

## RESULT 11

C38302  
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 23-Feb-1997  
C:Accession: C38302  
R:Kageyama, T.; Tanabe, K.; Koiwai, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu  
A:Reference number: A38302; MUID:91009127  
A:Accession: C38302  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.2%; Score 296; DB 2; Length 387;  
Best Local Similarity 26.9%; Pred. No. 1.3e-16;  
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

QY 30 YYVEMTVGSPPTNLINLVDTGSSNFAYGAAPHPF-----LHRYQRLSSTYRDLRK 82  
Db 75 YFGTISIGTTPQDFTVIFDFTGSSNLWV---PSTYCSSLACALHRRFPEDSDSYQGTSET 131



Qy	30	YVVEWTVGSPPTUNILNVDTGSSNFVAAPHPE-----LHRYORQLSSTYRDLRK 82
Db	75	YFGTISITGTPPDFTVIFDTGSSLNW----PSTYCSSLACALHKRFNPEDSSYTQTSET 131
Qy	83	VYPPTYOGKWEGLGTDLVSIPHPGNVTVRANIAATIESDKPF-----INGSNWE 132
Db	132	LSITYTGSGMTGILGYDTV-----KVGSIEDTNOIFGLSKTEPLGLTFELFAPFD 1797
Qy	133	GILGLAETARPPDSLEPFDFSLVKOTHV--PNULFSLOLCGAGFPPLNQSEVLASVGSGMI 191
Db	180	GILGLAVPSTISSDAT--PVFDNNWNNEGLYSQDLFSVYLSDD-----EKGLVM 227
Qy	192	IGGDHSLYTGSLWYTPIRREYYEVIIVRVINGOQDKM--DCKEYNYDKSVSDSGTTN 249
Db	228	FGGIDSSVYTGSLNMWPVSVEGYWQIITMDSVISINGETIACADSC-----QAIVDTCTSL 281
Qy	250	LRLPKKVPFAAVKSIKAASSTEKPDPGEFWIGEQLV-CWQAGTTPWNTFPVISLYLMGEVT 308
Db	282	LTGP-----TSAISNIQSYGASK----NLLGENVISCASIDLSPDIVF----- 321
Qy	309	NQSFRIITLPQYLPRVEDVATSODCCYKPAISOSSGT-----VMGAIMEGFYVVED 365
Db	322	-----TINGIQPLPASAYILTKDEDDCTCGLEGMNVDVTYGELWIILGDVFIRQYFTVED 3757
Qy	363	RARRKRIGFAVS 373
Db	376	RANNQLGLAAA 386
RESULT 15		
S19684		
pepsin A (EC 3.4.23.1) 2/3 precursor - Japanese macaque		
N;Alternate names: pepsinogen A isozyme 2/3		
C;Species: Macaca fuscata (Japanese macaque)		
C;Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1998		
C;Accession: S19684; S16064		
R;Kagayama, T.; Tanabe, K.; Koiwai, O.		
Eur J Biochem. 202, 205-215, 1991		
A;Title: Development-dependent expression of isozymogens of monkey pepsinoc		
A;Reference number: S19681; MUID: I2037645		
A;Accession: S19684		
A;Molecule type: mRNA		
A;Residues: 1-388 <KAG>		
A;Cross-references: EMBL:X59755; NID:g38068; PIDN:CAA42427.1; PID:g38069		
A;Note: parts of sequence, including amino ends of pepsinogen and activat		
C;Comment: It could not be determined if this sequence represents isozyme		
in by 4 months of age.		
C;Comment: Although two-step activation is observed, activation is predom		
C;Superfamily: pepsin		
C;Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein;		
F;1-15/Domain: signal sequence #3 status predicted &ltSIG>		
F;16-388/Product: pepsinogen A 2/3 #status experimental &ltPPT>		
F;16-62/Domain: activation peptide #status experimental &ltAPT>		
F;63-388/Product: pepsin A 2/3 #status experimental &ltBNZ>		
F;40-41/Cleavage site: Asp-Phe (pepsin) #status experimental		
F;62-63/Cleavage site: Leu-Ile (pepsin) #status experimental		
F;94,277/Active site: Asp #status predicted		
F;107-112,268-272,311-344/bisulfide bonds: #status predicted		
F;130/Binding site: phosphate (Ser) (covalent) #status predicted		

	Query Match	12.0%	Score 289.5;	DB 1;	Length 388;
	Best Local Similarity	27.0%;	Pred. No. 4.6e-16;		
	Matches 103;	Conservative 66;	Mismatches 141;	Indels 71;	Gaps
Qy	3 DEEPEPCRRGSFFVEMVDNLKKGSGQGYVVEMTVGSSPPQTILNILDVTGSSNFVAAGAAPH	62	:::   :	: :  :   :	:
Db	64 DEQPLE-----NYLDM-----EVFGTIGCTPAQFTVFDTGSSNLWPVSVC	108	:	:	:
Qy	63 FL----HRYQRQLSSTYRDLRKGVYPDYTGCKWEGELGTDLNSIPHGPNVTVRNTAAI	118	:	:	:
Db	109 SLACTNNHRNPQDSSTYSQTSNGVSYITYGTSMTGLGYDTGV--GGISDPTNQIFGL	165	:	:	:



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 17.62 Seconds

(without alignments)  
1002.050 Million cell updates/sec

Title: US-09-724-571-43

Perfect score: 2419

Sequence: 1 ETDEPPEPGRGSGFVEMVD.....CLRLRQHQDDFADDISLLK 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2419	100.0	501	1	BACE_HUMAN
2	2392	98.9	501	1	BACE_RAT
3	2390	98.8	501	1	BACE_MOUSE
4	1171.5	48.4	518	1	BAE2_HUMAN
5	327	13.5	324	1	PEP1_GADMO
6	314.5	13.0	350	1	CATD_BOVIN
7	306	12.6	387	1	PEP1_RABIT
8	302.5	12.5	388	1	PEP4_MACFU
9	302	12.5	367	1	PEPA_CHICK
10	299.5	12.4	396	1	CATE_HUMAN
11	298.5	12.3	383	1	PEPE_CHICK
12	296	12.2	387	1	PEP2_RABIT
13	296	12.2	391	1	CATE_CAVPO
14	295.5	12.2	412	1	CATD_HUMAN
15	295	12.2	387	1	PEP4_RABIT
16	289.5	12.0	388	1	PEP2_MACFU
17	287	11.9	407	1	CATD_RAT
18	286	11.8	387	1	PEP3_RABIT
19	285.5	11.8	398	1	CATE_RAT
20	285	11.8	398	1	CATD_CHICK
21	283.5	11.7	388	1	PEPA_MACMU
22	282.5	11.7	388	1	PEPA_HUMAN
23	281.5	11.6	381	1	CHYM_SHEEP
24	281.5	11.6	388	1	PEP1_MACFU
25	279.5	11.6	410	1	CATD_MOUSE
26	278.5	11.5	386	1	PEPA_PIG
27	278	11.5	387	1	PEPA_CALJA
28	277.5	11.5	396	1	CATE_RABIT
29	277.5	11.5	397	1	CATE_MOUSE
30	276.5	11.4	419	1	CARV_CANAL
31	273.5	11.3	381	1	CHYM_BOVIN
32	273	11.3	388	1	PEP4_BOVIN
33	270.5	11.2	377	1	PEP2_MACFU

34	270	11.2	376	1	PAG2_BOVIN	Q28057 bos taurus
35	267.5	11.1	381	1	CHYM_CALJA	Q9n2d2 callithrix
36	267	11.0	365	1	CATD_SHEEP	Q9mzs8 ovis aries
37	266	11.0	405	1	CARP_YEAST	P07267 saccharomyc
38	264	10.9	396	1	CARP_NEUCR	Q01294 neurospora
39	263.5	10.9	388	1	PEPC_CALJA	Q9n2d3 callithrix
40	262.5	10.9	394	1	PEPC_CAVPO	Q64411 cavia porce
41	261.5	10.8	387	1	ASPP_AEDAE	Q03168 aedes aegypt
42	261.5	10.8	388	1	PEPC_HUMAN	P20142 homo sapien
43	261	10.8	388	1	PAG_HORSE	Q28389 equus caball
44	260	10.7	345	1	CATD_PIG	P00795 sus scrofa
45	255	10.5	496	1	ASPR_ORYSA	P42211 oryza sativ

## ALIGNMENTS

RESULT 1

ID BACE\_HUMAN STANDARD; PRT; 501 AA.

AC P56817; Q9UJT5;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)

DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2) (Memapsin-2).

DE BACE OR BACE1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM A).

RC TISSUE=Brain;

RX MEDLINE=20002972; PubMed=10531052;

RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A., Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y., Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A., Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F., Treanor J., Rogers G., Citron M.;

RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";

RT Science 286:735-741(1999).

RL [2]

RP SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND CHARACTERIZATION.

RC TISSUE=Brain;

RX MEDLINE=20057171; PubMed=10591214;

RA Sinha S., Anderson J.P., Barbours R., Basl G.S., Caccavello R., Davis D., Doan M., Dovey H.F., Frigon N., Hong J., Jacobson-Croak K., Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H., Tatsuono G., Tung J., Schenk D., Seubert P., Suomensaar S.M., Wang S., Walker D., Zhao J., McConlogue L., Varghese J.;

RA "Purification and cloning of amyloid precursor protein beta-secretase from human brain.";

RT Nature 402:537-540(1999).

RL [3]

RP SEQUENCE FROM N.A. (ISOFORM A).

RX MEDLINE=20057170; PubMed=10591213;

RA Van R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M., Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B., Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;

RA "Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity.";

RT Nature 402:533-537(1999).

RL [4]

RP SEQUENCE FROM N.A. (ISOFORM A).

RX MEDLINE=20030186; PubMed=10561122;

RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D., Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M., Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;

\*Identification of a novel aspartic proteinase (Asp 2) as beta-secretase.\*;  
 RL Mol. Cell. Neurosci. 14:419-427(1999).

[5]  
 RN SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Brain, and Pancreas;  
 RA Michel B., De Pietri Tonelli D., Zaccchetti D., Keller P.;  
 RT "New beta-site APP cleaving enzyme isoform (BACE-1B) obtained from  
 human brain and pancreas.\*;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

[6]  
 RN SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.  
 RP MEDLINE=20144060; PubMed=10677483;  
 RX Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;

RA "Human aspartic protease memapsin 2, cleaves the beta-secretase site of  
 beta-amyloid precursor protein.\*;  
 RL Prog. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).

CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; A/BACE-1A (shown here) and  
 CC B/BACE-1B; are produced by alternative splicing.

CC -!- TISSUE SPECIFICITY: BRAIN.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL; AF190725; AAF04142.1; -

DR EMBL; AF201468; AAF18982.1; -

DR EMBL; AF200343; AAF17079.1; -

DR EMBL; AF204943; AAF26367.1; -

DR EMBL; AF338816; AAF38374.1; -

DR EMBL; AF200193; AAF13715.1; -

DR HSSP; P32329; LYPS.

DR MIM; 604252; -

DR InterPro: IPR001969; Asp.protease.

DR InterPro: IPR001461; pepsin.

DR Pfam; PF00026; asp. 3.

DR PRINTS; PR00792; PEPSIN.

DR PROSITE; PS00141; ASP\_PROTEASE; 1.

KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;

FT SIGNAL; Alternative splicing.

FT SIGNAL 1 21 POTENTIAL.

FT PROPEP 22 45

FT CHAIN 46 501 BETA-SECRETASE.

FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 458 478 POTENTIAL.

FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).

FT ACT\_SITE 93 93 BY SIMILARITY.

FT ACT\_SITE 289 289 BY SIMILARITY.

FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPIC 190 214 MISSING (IN ISOFORM B).

FT SEQUENCE 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;

Query Match

Best Local Similarity 100.0%; Score 2419; DB 1; Length 501;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEEPEERGRGSFVEMVDNLGRKSGQGYVYEMTVGSPQTNLNILDVTGSSNFVGAAP 60  
 DB |||||  
 DB 46 ETDEEPEERGRGSFVEMVDNLGRKSGQGYVYEMTVGSPQTNLNILDVTGSSNFVGAAP 105  
 QY 61 HPFLHRYQRLSSTYRDLRKGVVYPYTGKWEGLGTDLVSPHGPNTVVRANIAAITE 120  
 DB |||||  
 DB 106 HPFLHRYQRLSSTYRDLRKGVVYPYTGKWEGLGTDLVSPHGPNTVVRANIAAITE 165  
 QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLSPFFDSLVKQTHVNLFSQLCGAGFPPLNQS 180  
 DB |||||  
 DB 166 SDKFFINGSNWEGILGLAYAEIARPDSDLSPFFDSLVKQTHVNLFSQLCGAGFPPLNQS 225  
 QY 181 EVLASVGGSMIIGDHSILGSLWTPYIRREWYEVIIIVRVEINGODLKMCKEYNDK 240  
 DB |||||  
 DB 226 EVLASVGGSMIIGDHSILGSLWTPYIRREWYEVIIIVRVEINGODLKMCKEYNDK 285  
 QY 241 SIYDSGTNLRPKVFEAAVKSIAASSTKFPDGFGLGEQLVLCVQAGTTPWNIFPVIS 300  
 DB |||||  
 DB 286 SIYDSGTNLRPKVFEAAVKSIAASSTKFPDGFGLGEQLVLCVQAGTTPWNIFPVIS 345  
 QY 301 LYLMGEVTNQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVV 360  
 DB |||||  
 DB 346 LYLMGEVTNQSFRITILPQOYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVV 405  
 QY 361 FDRARKRIGFAVACHVHDEFRTAAVEGPFVTIDMDCGYNIPQTDSTLTMTIAYVMAAI 420  
 DB |||||  
 DB 406 FDRARKRIGFAVACHVHDEFRTAAVEGPFVTIDMDCGYNIPQTDSTLTMTIAYVMAAI 465  
 QY 421 CALFMPLCLMVCQWRCRLCRQHQHDDFADDISLLK 456  
 DB |||||  
 DB 466 CALFMPLCLMVCQWRCRLCRQHQHDDFADDISLLK 501

# RESULT 2

BACKE-RAT  
 ID BACKE-RAT STANDARD; PRT; 501 AA.

AC P56819;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)

DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl

DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)

DE (Memapsin-2).

GN BACE.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20002972; PubMed=10531052;

RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,

Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,

Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,

Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,

Treanor J., Rogers G., Citron M.;

RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by

the transmembrane aspartic protease BACE.\*;

Science 286:735-741(1999).

RL -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE

CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF

CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,

CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED

CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL

CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY

CC SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF190727; AAF04144.1; -.  
 DR HSP; P32329; 1YPS.  
 DR InterPro; IPR001969; Asp.protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp.3.  
 DR PRINTS; PR00792; PEPSIN.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
 KW SIGNAL.  
 FT PROPEP 1 21 POTENTIAL.  
 FT CHAIN 22 45 POTENTIAL.  
 FT DOMAIN 46 501 BETA-SECRETASE.  
 FT TRANSMEM 22 457 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 458 478 POTENTIAL.  
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 289 289 BY SIMILARITY.  
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 501 AA; 55806 MW; 24B445BC8B87DE3 CRC64;

Query Match 98.9%; Score 2392; DB 1; Length 501;  
 Best Local Similarity 98.2%; Pred. No. 9.2e-191;  
 Matches 446; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ETDEPEPRGRGSGFVEMVNLKSGGGYVEMTVGSPPTLILYDVTGSSNFAVGAAP 60  
 Db 46 ETDEPEPRGRGSGFVEMVNLKSGGGYVEMTVGSPPTLILYDVTGSSNFAVGAAP 105  
 QY 61 HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGNVTVRANIAITE 120  
 Db 106 HPFLHRYQRLSSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGNVTVRANIAITE 165  
 QY 121 SDRFFINGNWEGILGLAYAEIARPDLSLEPFFDLSLVKQTHVPLFSLQCGAGFPLNQS 180  
 Db 166 SDRFFINGNWEGILGLAYAEIARPDLSLEPFFDLSLVKQTHVPLFSLQCGAGFPLNQ 225  
 QY 181 EVLASVGSMLIGIDHSYLTGSLWYPIRREWYEVILVIRVEINGQDLKMDCKEYNDK 240  
 Db 226 EALASVGSMLIGIDHSYLTGSLWYPIRREWYEVILVIRVEINGQDLKMDCKEYNDK 285  
 QY 241 SIVDSGTTNLRPKKVFEEAAVKSIKAASSTEKFPDGFGLGEQLVQWAGTTPWNIFPVIS 300  
 Db 286 SIVDSGTTNLRPKKVFEEAAVKSIKAASSTEKFPDGFGLGEQLVQWAGTTPWNIFPVIS 345  
 QY 301 LYLMEVTNOSFRITILPOQLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGYVYV 360  
 Db 346 LYLMEVTNOSFRITILPOQLRPVEDVATSDQCYKFAISQSSTGTVMGAVIMEGYVYV 405  
 QY 361 FDRARKRIGFAVSACHVHDFRTAAVGGPVFTLMDCEGYNIPQTDSTLMTIAYVMAAI 420  
 Db 406 FDRARKRIGFAVSACHVHDFRTAAVGGPVFTADMECCGYNIFQTDSTLMTIAYVMAAI 465  
 QY 421 CALFMPLCLMVQWCRCLRCLRHQDHDFAADISLLK 456  
 Db 466 CALFMPLCLMVQWCRCLRCLRHQDHDFAADISLLK 501

RESULT 3

ID BACE\_MOUSE

AC P56818; STANDARD; PRT; 501 AA.

DT 30-MAY-2000 (Rel. 39, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
 DE (beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
 DE protease 2) (Asp 2) (Membrane-associated aspartic protease 2)  
 DE (Memapsin-2).  
 DE GN BACE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20002972; PubMed=10531052;  
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
 RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
 RA Treanor J., Rogers G., Citron M.;  
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
 the transmembrane aspartic protease BACE";  
 RL Science 286:735-741(1999).  
 RN [2]  
 RP REVISIONS TO 6 AND 81-87.  
 RA Bennett B.D., Vassar R., Citron M.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057170; PubMed=10591213;  
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,  
 RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
 RA Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;  
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease  
 beta-secretase activity";  
 RL Nature 402:533-537(1999).  
 CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP.  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: BRAIN.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF190726; AAF04143.2; -.  
 CC EMBL; AF200346; AAF17082.1; -.  
 CC HSP; P56272; 1AM5  
 CC MGD; MGI:1346542; Bace.  
 DR InterPro; IPR001969; Asp.protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp.1.  
 DR PRINTS; PR00792; PEPSIN.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
 KW SIGNAL.  
 FT PROPEP 1 21 POTENTIAL.  
 FT CHAIN 22 45 POTENTIAL.  
 FT CHAIN 46 501 BETA-SECRETASE.  
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 458 478 POTENTIAL.  
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 289 289 BY SIMILARITY.

FT	CARBOHYD	153	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	172	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	223	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	354	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	501 AA; 55747 MW; C085A013145E474E CRC64;	
Query Match 98.8%; Score 2390; DB 1; Length 501;			
Best Local Similarity 98.2%; Pred. No. 1.3e-190;			
Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;			
QY	1	ETDEEPEEGRGRSFVEMVNLKRGKSGQGYVEMTVGSPQTNLILVDTGSSNFVGAAP	60
DB	46	ETDEESEPGRGRSFVEMVNLKRGKSGQGYVEMTVGSPQTNLILVDTGSSNFVGAAP	105
QY	61	HPFLHRYQRLSTYRDLRKGVYVPTQCKWEGELGTLVSIHPGPNVTVRANIAAITE	120
DB	106	HPFLHRYQRLSTYRDLRKGVYVPTQCKWEGELGTLVSIHPGPNVTVRANIAAITE	165
QY	121	SDKFFINGNWEGILGLAYAEIARPDLSLEPFFDSLVKQTHVPLMFLSLQLCGAGFPLNQS	180
DB	166	SDKFFINGNWEGILGLAYAEIARPDLSLEPFFDSLVKQTHVPLMFLSLQLCGAGFPLNQT	225
QY	181	EVLASVGSMIIGIDHSLYTGSWYTPIRREWYVEIIVRVEINGODLKMCKEYNKYD	240
DB	226	EALASVGSMIIGIDHSLYTGSWYTPIRREWYVEIIVRVEINGODLKMCKEYNKYD	285
QY	241	SIYDSGTTNLRPKKFEAAVKSTKAASSTKPDGFWLGEQLVCGWAGTTPNNIPPVIS	300
DB	286	SIYDSGTTNLRPKKFEAAVKSTKAASSTKPDGFWLGEQLVCGWAGTTPNNIPPVIS	345
QY	301	LYLMGEVTNOSFRITILPOOLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVV	360
DB	346	LYLMGEVTNOSFRITILPOOLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVV	405
QY	361	FDRARKRIGFAVSACHVHDEFTAAVGPFWLDMEDCGYNIPQTDSTLMTIAYVMAAI	420
DB	406	FDRARKRIGFAVSACHVHDEFTAAVGPFWLDMEDCGYNIPQTDSTLMTIAYVMAAI	465
QY	421	CALFMLPCLMVCQWRCLRLRQOHHDFADDISLLK	456
DB	466	CALFMLPCLMVCQWRCLRLRQOHHDFADDISLLK	501
RESULT 4			
BAE2_HUMAN	STANDARD;	PRT;	518 AA.
ID	BAE2_HUMAN	Q9Y520; Q9UT6;	
AC	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving		
DE	enzyme 2) (Aspartyl protease 1) (ASP 1) (ASP1) (Membrane-associated		
DE	aspartic protease 1) (Memapsin-1).		
GN	BACE2 OR ASP21.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20057170; PubMed=10591213;		
RA	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,		
RA	Brashler J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,		
RA	Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;		
PT	"Membrane-anchored aspartyl protease with Alzheimer's disease		
RT	beta-secretase activity."		
RL	Nature 402:533-537(1999).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUP=Bone marrow;		
RA	Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,		
RA	Gliese K.;		

RT	"Identification of a novel aspartic-like protease differentially		
RT	expressed in human breast cancer cell lines."		
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RP	[3]		
RP	SEQUENCE FROM N.A.		
RA	Accarino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;		
RT	"Cloning of a gene from chromosome 21 down region encoding a potential		
RT	transmembrane aspartyl protease."		
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.		
RP	[4]		
RP	SEQUENCE FROM N.A.		
RA	Solaus A., Estivill X., de la Luna S.;		
RT	"Cloning of a novel mammalian aspartyl protease."		
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		
RP	[5]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20030166; PubMed=10561122;		
RA	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,		
RA	Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,		
RA	Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;		
RT	"Identification of a novel aspartic proteinase (Asp 2) as		
RT	beta-secretase."		
RL	Mol. Cell. Neurosci. 14:419-427(1999).		
RP	[6]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20144060; PubMed=10677483;		
RA	Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;		
RT	"Human aspartic protease memapsin 2 cleaves the beta-secretase site of		
RT	beta-amyloid precursor protein."		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).		
RP	[7]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20289799; PubMed=10830953;		
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,		
RA	Park H.-S., Toyoda A., Ishii K., Tokoki Y., Choi D.-K., Soeda E.,		
RA	Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,		
RA	Manzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,		
RA	Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,		
RA	Rosenthal A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,		
RA	Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,		
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,		
RA	Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,		
RA	Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,		
RA	Lehrach H., Reinhardt R., Yaspo M.-L.;		
RT	"The DNA sequence of human chromosome 21."		
RL	Nature 405:311-319(2000).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE		
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL; AF200342; AAF17078.1; -		
DR	EMBL; AF117892; AAD45240.1; -		
DR	EMBL; AF050171; AAD45963.1; -		
DR	EMBL; AF178532; AAF29494.1; -		
DR	EMBL; AF204944; AAF26368.1; -		
DR	EMBL; AF200192; AAF13714.1; -		
DR	EMBL; AL163284; CAB90458.1; -		
DR	EMBL; AL163285; CAB90354.1; -		
DR	HSSP; P00797; 2REN.		
DR	MM; 605668; -		
DR	InterPro; IPR001969; Asp_protease.		
DR	InterPro; IPR001461; Pepsin.		
DR	Pfam; PF00026; asp; 3.		
DR	PRINTS; PR00792; PEPSIN.		



```
DR PROSITE; PS00141; ASP_PROTEASE; 2.  
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
FT SIGNAL. 1 20 POTENTIAL.  
FT PROPEP 21 ? POTENTIAL.  
FT CHAIN ? 518 BETA SECRETASE 2.  
FT DOMAIN 21 473 EXTRACELLULAR (POTENTIAL).  
FT TRANSEM 474 494 POTENTIAL.  
FT DOMAIN 495 518 CYTOPLASMIC (POTENTIAL).  
FT ACT_SITE 110 110 BY SIMILARITY.  
FT ACT_SITE 303 303 BY SIMILARITY.  
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 36 36 A -> T (IN REF. 6).  
SQ SEQUENCE 518 AA; 56180 MW; 2E903150823760D3 CRC64;  
  
Query Match 48.4%; Score 1171.5; DB 1; Length 518;  
Best Local Similarity 49.6%; Pred. No. 1.6e-89;  
Matches 222; Conservative 80; Mismatches 139; Indels 7; Gaps 4;  
  
QY 9 PGRGSEVEMVNDLRKSGGQYVYVETVGSPPOTLNLIVDTGSSNFVAGRAPHPFLHRY 68  
DB 71 PAGAANFLAMVNDLQGGSGRGYILEMIGTPPKLQLILVDTGSSNFVAGTPHSHYIDTF 130  
QY 69 QRLSSYRDLRKGVVYPTQGWEGELGTDLSVPHGPNVTVRANIAAITESDKFFING 128  
DB 131 DTERSSTYRSGFDVTKYTGGSWTFVGEDLVTPKGFNTSFLVNATIFESNFFLP 190  
QY 129 SNWEGILGLAYEARTDDLSLEPPFDSLVKQTHVNFSLQLCGAGPPLNQSEVLASVG 188  
DB 191 IKWNGILGLAYATLAKPSSSLETFFDSLVQTQANIPNPFVSMQCGAGLPVAGS---GTNG 247  
QY 189 SMIIIGGDHSLYTGSLWATPTIRREWYEVLIIVRVEINGDLKMDCKEYNYDKSIVDSGTT 248  
DB 248 SLVGGLEPSLYKGDWYTIKEWYIQIELLEIGQSLNLDREYNADKAIVDSGTT 307  
QY 249 NLRPKVFEAAVKSIAKASTKFPDGLGEQLVCWQAGTTPWNIFFPVSILYLMGEVT 308  
DB 308 LLRLPQKVFDAVVEAVARASLIPESDGFSGTGSQACWNTSETPWSYFPKISIVLRDENS 367  
QY 309 NQSPRIITLPOQLRPVEDVATSDCCYKFAISQSSTGTWNGAVIMEGFYVVDRAKRI 368  
DB 368 SRSFRITLPQIXQPMGAGLNY-ECYRFGISPSTNALVIGATVMEGFYVVDRAQKV 426  
QY 369 GFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAVMAAIC-ALPMLP 427  
DB 427 GFAASPCAETAGAAVSEISGPFSTEDVASNCVPAQSUSEPILWIVSYALMSVCGAILLV 486  
QY 428 LCLMVCQWCLRLCRLQHQHDDFADDISLL 455  
DB 487 IVLLLLPFCRQ--RPRDPEVWDESSL 512
```

```
RESULT 5  
PEPL_GADMO  
ID PEPL_GADMO STANDARD; PRT; 324 AA.  
AC P56272;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Pepsin IIB (EC 3.4.23.-).  
OS Gadus morhua (Atlantic cod).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;  
OC Gadus.  
OX NCBI_TaxID=8049;  
RN [1]  
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.  
RC TISSUE=Stomach;  
RA Karlsten S., Hough E., Olsen R.L.;  
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
```

```
RT cod (Gadus morhua).";  
RL Acta Crystallogr. D 54:32-46(1998).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR PDB; IAMS; 24-DEC-97.  
DR InterPro; IPR001969; Asp.protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF000026; asp.1.  
DR PRINGS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP_PROTEASE; 2.  
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
3D-structure. 32 32 BY SIMILARITY.  
FT ACT_SITE 214 214 BY SIMILARITY.  
FT ACT_SITE 45 50 BY SIMILARITY.  
FT DISULFID 206 209 BY SIMILARITY.  
FT DISULFID 247 280 BY SIMILARITY.  
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;  
  
Query Match 13.5%; Score 327; DB 1; Length 324;  
Best Local Similarity 27.9%; Pred. No. 9.1e-20;  
Matches 104; Conservative 66; Mismatches 137; Indels 66; Gaps 15;  
  
QY 18 MVDNLRKSGQYVYVETVGSPPOTLNLIVDTGSSNFVAG---AAPHPFLHRYQRLS 73  
DB 2 VTEQMKNEADTEYIGVISIGTPPESFKVIFDTGSSNLWVSSSHCSQAQCSNNHKKFPRQS 61  
QY 74 STYRDLRKGVVYPTQGWEGELGTDLSVPHG---PNVTVRANIAAITESDKFFINGSNW 131  
DB 62 STYVETKTVDLTVGTGMRGILQDQTVSVGGSDPNQELG---ESQTEPGPQA-AAPF 117  
QY 132 EGILGLAYAIARDDLSLEPPFDSLVKQTHV-PNLSQLCGAGPPLNQSEVLASVGSM 190  
DB 118 DGIILGLAYATLAKPSSSLETFFDSLVQTQANIPNPFVSMQCGAGLPVAGS---ANGSEVM 166  
QY 191 IIGGDHSLYTGSLWATPTIRREWYEVLIIVRVEINGDLKMD-CKEYNYDKSIVDSGTTN 249  
DB 167 -LGGVDSNHTYGSIHWPVTAEKYQWQVALDGIPTVNGQTAACEG-----QAIVDTGTSK 219  
QY 250 LRLPKVFEAAVKSIAKASTKFPDGLGEQLVCWQAGTTPWNIFFPVSILYLMGEVTN 309  
DB 220 IVAPVSALANIMKDIGASEN---QCEMGN---CASVQSLPDITF-----TI 260  
QY 310 QSPRIITLPOQLRPVEDVATSDCCYKFAISQSSTGT-----VNGAVIMEGFYVVF 361  
DB 261 NGVKQPLPPSAYIEGDAQFCTS-----GLGSSGVPSTSELWIFGDFVLRNYTYI 311  
QY 362 DRARKRIGFAVSA 374  
DB 312 DRTNKKVGFAPAA 324  
  
RESULT 6  
CATD_BOVIN  
ID CATD_BOVIN STANDARD; PRT; 390 AA.  
AC P80209; Q9TS27;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cathepsin D precursor (EC 3.4.23.5).  
GN CTSd.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI_TaxID=9913;  
RN [1]  
RP SEQUENCE OF 1-48.  
RC TISSUE=Milk;  
RX MEDLINE=93202276; PubMed=8454061;  
RA Larsen L.B., Boisen A., Petersen T.E.;  
RT "Procathepsin D cannot autoactivate to cathepsin D at acid pH.";
```

```
RL FEBS Lett. 319:54-58(1993).
RN [2]
RP SEQUENCE OF 45-390, AND X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RC TISSUE=Liver;
RA MEDLINE=93223670; PubMed=8467789;
RX Metcalf P., Fusek M.;
RT "Two crystal structures for cathepsin D: the lysosomal targeting
RL signal and active site.";
RL EMBO J. 12:1293-1302(1993).
CC -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
CC PROTEIN BREAKDOWN.
CC -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
CC that of pepsin A. Does not cleave the 4-Gln-|-His-5 bond in B
CC chain of insulin.
CC -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; S32383; S32383.
DR PIR; S37419; S37419.
DR HSSP; P07339; 1LYB.
DR MEROPS; A01.009; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Zymogen.
FT PROPEP 1 44
FT CHAIN 45 390
FT ACT_SITE 77 77
FT ACT_SITE 273 273
FT DISULFID 71 140
FT DISULFID 90 97
FT DISULFID 264 268
FT DISULFID 307 344
FT CARBOHYD 114 114
FT CARBOHYD 241 241
FT SEQUENCE 390 AA; 42488 MW; 5B39AA1C33C48D35 CRC64;
Query Match 13.0%; Score 314.5; DB 1; Length 390;
Best Local Similarity 28.0%; Pred. No. 1.3e-18;
Matches 113; Conservative 72; Mismatches 128; Indels 91; Gaps 21;
QY 8 EPG-RRGSFVEMVDNLGRKSGQYVEMTVGSPQTNLILVDTGSSNFAYGAAPHPFL-- 64
DB 39 EPAVRQGFPELLKNYDAQ---YYGEIGIGTPQCFTVVFDGTSANLWVPSIHCKLLDI 95
QY 65 ----HRYQRLSTYRDLRKGVY--VPYTGKKEGELGTDLVSIPIGPN-----VTVR 112
DB 96 ACWTHRYKNSDKSSTY--VKNGTTFDIHYGSGSLGSLSDQTVSVPCNPSSSPGGVTVQ 153
QY 113 ANI--AAITESDKFINGSNWEGILGLAYAEIARPDOSLEPFDSLAVKQTHV-PNLFSIQ 169
DB 154 RQTFGEAIKQGVVFI-AAKFDILGMAYPRIS---VNNVLFPDNLMOQKLVDRKNFS-- 208
QY 170 LCGAGFPLNQSEVLASVGGSMIIGIDHSILYTGSLWTPTIRREWYVYIIVRVEINGQDL 229
DB 209 -----FFLNR-DPKAQPGLMGLGTDKSKYRGSLMFHNTVQAYWQIHMQLDV-GSSL 261
QY 230 KMDCKEYNYKSDVSTGNTNLRPKKVFEEAAVSIKAASTKEKPPDGFWEQLV-CWQA 288
DB 262 TV-CK--GGCEAIVDTGTSLVGPVEEVREIQAIGAVPLIQ-----GEYMIPEKV 310
QY 289 GTTPNFPVLSLGMENVNQSRITILPQOYLRPEVDVATSDCCYKFAISOSSTGT- 347
DB 311 SS-----LPEVTVKLGG-----KDYALSPED-YALKVSAQETVC 344
QY 348 -----VMGAVIMEGFVYVDFDRARKRIGFAVSA 374
DB 345 LSGFMGMDIPPPGGLWILGDFVIGRYVTVFDRQNRVGLAEAA 388
```

```
RESULT 7
PEP1_RABIT STANDARD; PRT; 387 AA.
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin II-1 precursor (EC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; B38302; B38302.
DR HSSP; P00791; 1PSA.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59
FT CHAIN 60 387
FT MOD_RES 129 129
FT ACT_SITE 93 93
FT ACT_SITE 276 276
FT DISULFID 106 111
FT DISULFID 267 271
FT DISULFID 310 343
FT SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;
Query Match 12.6%; Score 306; DB 1; Length 387;
Best Local Similarity 27.1%; Pred. No. 6.4e-18;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;
QY 30 YVEMTVGSPQTNLILVDTGSSNFAYG----AAPHPHLHRYQRLSTYRDLRKGVYV 85
DB 75 YFGFISIGTPPEFTVIFDTGSSNLWVPSTYCCSLACFLHKRFNPDDSTFQATSETLSI 134
QY 86 PYTGCKEGLGTDLVSIPIGHPNVTVRANITAEISD---KFFINGSNWEGILGLAYAEI 142
DB 135 TYGTGSMTGILGYDTVKV---GNIEDINQIFGLSKTEPGITFLV--APFDGILGLATPSI 189
QY 143 ARPDOSLEPFDSLAVKQTHV-PNLFSIQLCGAGFPLNQSEVLASVGGSMIIGIDHSILY 201
DB 190 SASDAT--PFDNNWNEGLVSEDLFSVYLSNG-----EKGSVMVEGIDSSYTT 237
QY 202 GSLWYTPIRREWYVYIIVRVEINGQDLKM--DCKEYNYKSDVSTGNTNLRPKKVFEEA 259
DB 238 GSNLWVPVSEGYWQITMDSITINGETIACADSC-----QAVVDTGTSLLAGPTSAISK 291
```







```
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59235; AAA85369.1; -.
CC PIR; C38302; C38302.
CC HSP; P00790; 1PSN.
CC MEROPS; A01.001; -.
CC InterPro; IPR001969; Asp_protease.
CC Pfam; PF00026; asp; 1.
CC PRINTS; PR00792; PEPSIN.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PEPSIN II-2/3.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT ACT_SITE 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
FT SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;
CC -----
Query Match 12.2%; Score 296; DB 1; Length 387;
Best Local Similarity 26.9%; Pred. No. 4.3e-17;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;
QY 30 YVEMTVGSPPTLNILVDGTGSSNFAYGAAPHP-----LHRYQRLSSSTYDLRKG 82
D 30 YVEMTVGSPPTLNILVDGTGSSNFAYGAAPHP-----LHRYQRLSSSTYDLRKG 82
D 75 YFGTISGTGPPQDFVTFDGTSSNLWV--PSTYCSSLACALHKRFNEDSSTVQGTSET 131
QY 83 YVYPYTGKWEGLGTLVSIPIHGPNTVTVRANIAAITESSDKFFINGSNWEGILGLAYAEI 142
D 83 YVYPYTGKWEGLGTLVSIPIHGPNTVTVRANIAAITESSDKFFINGSNWEGILGLAYAEI 142
D 132 LSIYTGTSMTGILGTVKVGSTIEDNQIFGLSKTPEPSLTLF--APFDGILGLAYPSI 189
QY 143 ARPDLSLEPFDSLVAKTHV--PNLFSQLCGAGPLNQSEVLASVSGSMIIGGDHSLYT 201
D 143 ARPDLSLEPFDSLVAKTHV--PNLFSQLCGAGPLNQSEVLASVSGSMIIGGDHSLYT 201
D 190 SSDAT--PVFDNNWNEGLVSDQLFSVILSSDD-----EKGSLVFEGLDSSYYT 237
QY 202 GSLWYTPIRREWEYEVIVRVEINGODLKM--DCKEYNDKSIYDSGTTNLRPKKVFEEA 259
D 202 GSLWYTPIRREWEYEVIVRVEINGODLKM--DCKEYNDKSIYDSGTTNLRPKKVFEEA 259
D 238 GSLNWPVSVYEGYQWITMDSVISINGETIACADSC-----QAIVDTGTSLLTGP---TS 287
QY 260 AVSIIKAASSTKPEPDGFWGLGQVLV-CWQAGTTPNWIFPVVLSLIMGEVTVNQSFRIILP 318
D 260 AVSIIKAASSTKPEPDGFWGLGQVLV-CWQAGTTPNWIFPVVLSLIMGEVTVNQSFRIILP 318
D 288 AISNIQSYIGASK-----NLIGENVISCAIDSLDPIVF-----TING 325
QY 319 QOYLRPVEDVATSDCCYKFAISOSSTGT-----VMGAVIMEGFYVVFVRARKRIGFAV 372
D 319 QOYLRPVEDVATSDCCYKFAISOSSTGT-----VMGAVIMEGFYVVFVRARKRIGFAV 372
D 326 IQYPLPASAYILKEDDCTSCGLEGMNVDTYTGELWILGVDVFIROYFVFDVDRANQLGLAA 385
QY 373 S 373
D 386 A 386
RESULT 13
CATE_CAVPO STANDARD; PRT; 391 AA.
ID CATE_CAVPO
AC P25796;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Cathepsin E precursor (EC 3.4.23.34).
CTSE.
GN Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Mystricognathi; Cavidae; Cavia.
OX NCBI_Taxid=10141;
RN [1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92355614; PubMed=1644829;
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koizumi O.,
RA Tanji M., Yakabe E., Athuda S.B., Takahashi K.;
RT "Gastric procathepsin E and progastricsin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E.";
RL J. Biol. Chem. 267:16450-16459(1992).
RN [2]
RN SEQUENCE FROM N.A.
RX TISSUE=Gastric mucosa;
RX MEDLINE=96073637; PubMed=8540321;
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
RA Tanji M., Athuda S.B., Takahashi K.;
RT "Isolation, characterization, and structure of procathepsin E and
RT cathepsin E from the gastric mucosa of guinea pig.";
RL Adv. Exp. Med. Biol. 362:211-221(1995).
CC -!- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -!- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
CC specificity.
CC -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88653; AAA37052.1; -.
CC EMBL; S80547; AAB35844.1; -.
CC PIR; A43356; A43356.
CC HSP; P00794; 4CMS.
CC MEROPS; A01.010; -.
CC InterPro; IPR001969; Asp_protease.
CC InterPro; IPR001461; Pepsin.
CC Pfam; PF00026; asp; 1.
CC PRINTS; PR00792; PEPSIN.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 53 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 54 391 CATHEPSIN E.
FT ACT_SITE 92 92 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 56 56 INTERCHAIN (PROBABLE).
FT DISULFID 105 110 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 309 346 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC...)(POTENTIAL).
FT SEQUENCE 391 AA; 42132 MW; 78D216BF8CFDABD CRC64;
CC -----
Query Match 12.2%; Score 296; DB 1; Length 391;
Best Local Similarity 26.9%; Pred. No. 4.4e-17;
Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;
QY 30 YVEMTVGSPPTLNILVDGTGSSNFAYGA---APHFLHRYQRLSSSTYDLRKGYYV 85
D 30 YVEMTVGSPPTLNILVDGTGSSNFAYGA---APHFLHRYQRLSSSTYDLRKGYYV 85
D 74 YFGTISGTGPPQNFVTFDGTGSSNLWVPSVYCTSPACOTHPVFPVHPSJSTYREGVNSFSI 133
QY 86 PYTGKWEGLGTLVSIPIHGPNTVTVRANIAAITESSDKFFINGSNWEGILGLAYAEIARP 145
```

Db 134 QYGTSLGIIIGADQVSV-EGLTVVVGQFGEVQEPGKTEVH-AEFDGLIGLGYSLAA- 190  
Qy 146 DDLSEPFDSLQKPHVNLPSLQCLGAGFPLNQSEVLASVGGSMIIIGDHSLYTGSIM 205  
Db 191 -GGVTPVDFNMMAQ-----NLVALPM-----FSYMSNPGSGSELTFGGYDPFSHFGSLN 241  
Qy 206 YTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSGTNTLRLPKKVFEEAAVKSIK 265  
Db 242 WVPVTQKAYWQIALDGIQVG--DSVMFCSE--GCQAIYDTGTSLTGP-----PGKIKQLQ 293  
Qy 266 AASSTEKPFDFGLGEQLVQWQAGTPPNIPVISLYLMGEVYNOSFRI-----TILPOQ 320  
Db 294 EALGATYVDEGY-----SVQC-----ANLNMLDVT-----FIINGVVPYTLNPTA 333  
Qy 321 YLRPVEDVATSDQDCYKFAISQSSFG-----TWGAVIMEGFVVVEDRKR 367  
Db 334 Y--TLLDFVDMQVC-----STGFEGLEIQPPAGPLWILGDVFIHQFVAVFDGRGNR 383  
Qy 368 IGFA 371  
Db 384 VGLA 387

RESULT 14  
CATD\_HUMAN  
ID CATD\_HUMAN STANDARD; PRT; 412 AA.  
AC P07339;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Cathepsin D precursor (EC 3.4.23.5).  
OS CTSD.  
GN Homo sapiens (Human).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85270436; PubMed=3927292;  
RA Faust P.L., Kornfeld S., Chirgwin J.M.;  
RT "Cloning and sequence analysis of cDNA for human cathepsin D.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87231068; PubMed=3588310;  
RA Westley B.R., May F.E.B.;  
RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive  
RL human breast cancer cells.";  
RN Nucleic Acids Res. 15:3773-3786(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91299158; PubMed=2069717;  
RA Redeker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;  
RT "Molecular organization of the human cathepsin D gene.";  
RL DNA Cell Biol. 10:423-431(1991).  
RN [4]  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE=94085791; PubMed=8262386;  
RA May F.E., Smith D.J., Westley B.R.;  
RT "The human cathepsin D-encoding gene is transcribed from an estrogen-  
RL regulated and a constitutive start point.";  
RN Gene 134:277-282(1993).  
RN [5]  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE=95021301; PubMed=7935485;  
RA Augereau P., Miralles F., Cavailles V., Gaudelet C., Parker M.,  
RL Rochefort H.;  
RT "Characterization of the proximal estrogen-responsive element of  
RL human cathepsin D gene.";  
RN Mol. Endocrinol. 8:693-703(1994).  
RN [6]  
RP SEQUENCE OF 170-180.

RC TISSUE=Liver;  
RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,  
RA Pasquelli C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,  
RA Appel R.D., Hughes G.J.;  
RL Submitted (JUN-1992) to the SWISS-PROT data bank.  
RN [7]  
RP VARIANT VAL-58.  
RX MEDLINE=20179010; PubMed=10716266;  
RA Papasotiropoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,  
RA Maier W., Pauls J., Lautenschlager N., Heun R.;  
RT "A genetic variation of cathepsin D is a major risk factor for  
RL Alzheimer's disease.";  
RN Ann. Neurol. 47:399-403(2000).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).  
RC TISSUE=Spleen;  
RX MEDLINE=93223670; PubMed=8467789;  
RA Metcalfe P., Fusek M.;  
RT "Two crystal structures for cathepsin D: the lysosomal targeting  
RL signal and active site.";  
RN EMBO J. 12:1293-1302(1993).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RC TISSUE=Liver;  
RX MEDLINE=93342076; PubMed=8393577;  
RA Baldwin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,  
RA Cachau R.E., Collins J., Silva A.M., Erickson J.W.;  
RT "Crystal structures of native and inhibited forms of human cathepsin  
RL D: implications for lysosomal targeting and drug design.";  
RN Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).  
CC -!- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR  
CC PROTEIN BREAKDOWN.  
CC -!- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,  
CC that of pepsin A. Does not cleave the 4-Gln-1-His-5 bond in B  
CC chain of insulin.  
CC -!- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- POLYMORPHISM: The Val-58 allele is significantly overrepresented  
CC in demented patients (11.8%) compared with nondemented controls  
CC (4.9%). Carriers of the Val-58 allele have a 3.1-fold increased  
CC risk for developing AD than noncarriers.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M11233; AAB59529.1; -;  
DR EMBL; X05344; CAA28955.1; -;  
DR EMBL; M63138; AAA51922.1; -;  
DR EMBL; M63134; AAA51922.1; JOINED.  
DR EMBL; M63135; AAA51922.1; JOINED.  
DR EMBL; M63136; AAA51922.1; JOINED.  
DR EMBL; M63137; AAA51922.1; JOINED.  
DR EMBL; L12980; AAA16314.1; -;  
DR EMBL; S74689; AAD14156.1; -;  
DR EMBL; S52557; AAD13868.1; -;  
DR PIR; A25771; KKHUO.  
DR PDB; 1LYA; 3I-JAN-94.  
DR PDB; 1LYB; 3I-JAN-94.  
DR MEROPS; A01.009; -;  
DR SWISS-2DPAGE; P07339; HUMAN.  
DR Siena-2DPAGE; P07339; -;  
DR MIM; 116840; -;  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.

DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
KW Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Signal; Zymogen;  
KW Polymorphism; Alzheimer's disease; 3D-structure.  
FT SIGNAL 1 18  
FT PROPEP 19 64 ACTIVATION PEPTIDE.  
FT CHAIN 65 412 CATHEPSIN D.  
FT CHAIN 65 161 LIGHT CHAIN (PROBABLE).  
FT CHAIN 169 412 HEAVY CHAIN (PROBABLE).  
FT ACT\_SITE 97 97  
FT ACT\_SITE 295 295  
FT DISULFID 91 160  
FT DISULFID 110 117  
FT DISULFID 286 290  
FT DISULFID 329 366  
FT CARBOHYD 134 134  
FT CARBOHYD 263 263  
FT VARIANT 58 58  
FT  
FT STRAND 67 74  
FT TURN 75 77  
FT STRAND 78 85  
FT STRAND 86 89  
FT TURN 90 97  
FT STRAND 98 99  
FT TURN 103 107  
FT TURN 108 109  
FT TURN 112 113  
FT HELIX 115 118  
FT TURN 119 119  
FT STRAND 123 123  
FT HELIX 125 127  
FT TURN 129 130  
FT STRAND 132 141  
FT STRAND 146 158  
FT STRAND 172 184  
FT HELIX 188 192  
FT STRAND 197 200  
FT TURN 204 206  
FT TURN 208 210  
FT HELIX 214 220  
FT TURN 221 222  
FT STRAND 228 233  
FT STRAND 243 247  
FT TURN 248 248  
FT HELIX 252 254  
FT STRAND 255 263  
FT STRAND 267 267  
FT TURN 268 269  
FT STRAND 270 279  
FT TURN 280 281  
FT STRAND 284 285  
FT TURN 287 288  
FT STRAND 290 294  
FT TURN 296 297  
FT STRAND 298 298  
FT STRAND 301 303  
FT HELIX 305 315  
FT TURN 316 316  
FT STRAND 318 319  
FT TURN 322 323  
FT STRAND 325 328  
FT HELIX 329 334  
FT STRAND 338 342  
FT TURN 343 344  
FT STRAND 345 349  
FT HELIX 351 354  
FT STRAND 355 357  
FT TURN 359 368  
FT STRAND 365 368  
FT STRAND 370 372

N-LINKED (GLCNAC. . .).  
N-LINKED (GLCNAC. . .).  
A -> V (ASSOCIATED WITH INCREASED RISK IN  
AD; POSSIBLY INFLUENCES SECRETION AND  
INTRACELLULAR MATURATION; IN  
DBSNP:17571).  
/FtId=VAR\_011621.

Query Match 12.2%; Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. No. 5.2e-17;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;  
QY 30 YYVEMTVGSPPTLNILVDTGSSNFAYGAAPHPFL-----HRYYQRLSTYDRLRGV 83  
DB 79 YGEGIGTPTTQCFTVVDGSSNLWVPSIHCKLLDIACWIHHKYSNDSKSTYYKNGTSF 138  
QY 84 YVPTQKNGELGTDLVSP-----HGNVTVRANIAAITESDKPFINGNWEGI 134  
DB 139 DIHYGSGSLGYLSQDTSVPCQSASASALGVGVKVERQVGEATKQPGITFIKAFDGI 198  
QY 135 LGLAYAEIARPDSDLEPPFDLSLVKQTHV-PNLSIQLCGAGFPLNQSEVLASVGGSMIG 193  
DB 199 LGMAYPRIS--VNNVLPVFDNLMOQKLVQDNIFSYL-----SRPDAQPGGELMIG 248  
QY 194 GIDHSLYTGLWYTPIRREWYEVIIVRVEI-NGODLKMCKEYNDKSIYVDSGTTNRL 252  
DB 249 GTDSKYYKGSLSYLNWTRKAYQVHLQDQVEVASGLTL---CKE--GCEAIVDTGTSLMVG 303  
QY 253 PKVFEAAVKSIIKAASTETKFPDGFNLGEOLV-CWQAGTTTPWNIFFVISLYLMGEVNTQS 311  
DB 304 PVDEVRELQKAIGAVPLIQ-----GEYMIPCEKVST-----LPAITLKGK---KG 346  
QY 312 FRITILPOOYLRPVEDVATSDQDCYKFAISO-----SSTGTVMGAVIMEGFYVVFDRARK 366  
DB 347 YKLS--PETYLVKVSQAKTL--CLSGFMGMDIPPPSGPLWILGDVFGRYTYTFEDRNN 402  
QY 367 RIGFAVSA 374  
DB 403 RVGFAEAA 410

RESULT 15  
PEP4\_RABIT  
ID PEP4\_RABIT STANDARD; PRT; 387 AA.  
AC P28713;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Pepsin II-4 precursor (EC 3.4.23.1) (Pepsin A).  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91009127; PubMed=2129536;  
RA Kageyama T., Tanabe K., Koizumi O.;  
RT "Structure and development of rabbit pepsinogens. Stage-specific  
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and  
RT gene expression during development.";  
PL J. Biol. Chem. 265:17031-17038(1990).  
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
CC ALSO CLEAVED TO SOME EXTENT.  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE  
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.  
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
CC HORMONES AND RELATED SUBSTANCES.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR PIR; D38302; D38302.  
DR HSSP; P00790; 1PSN.

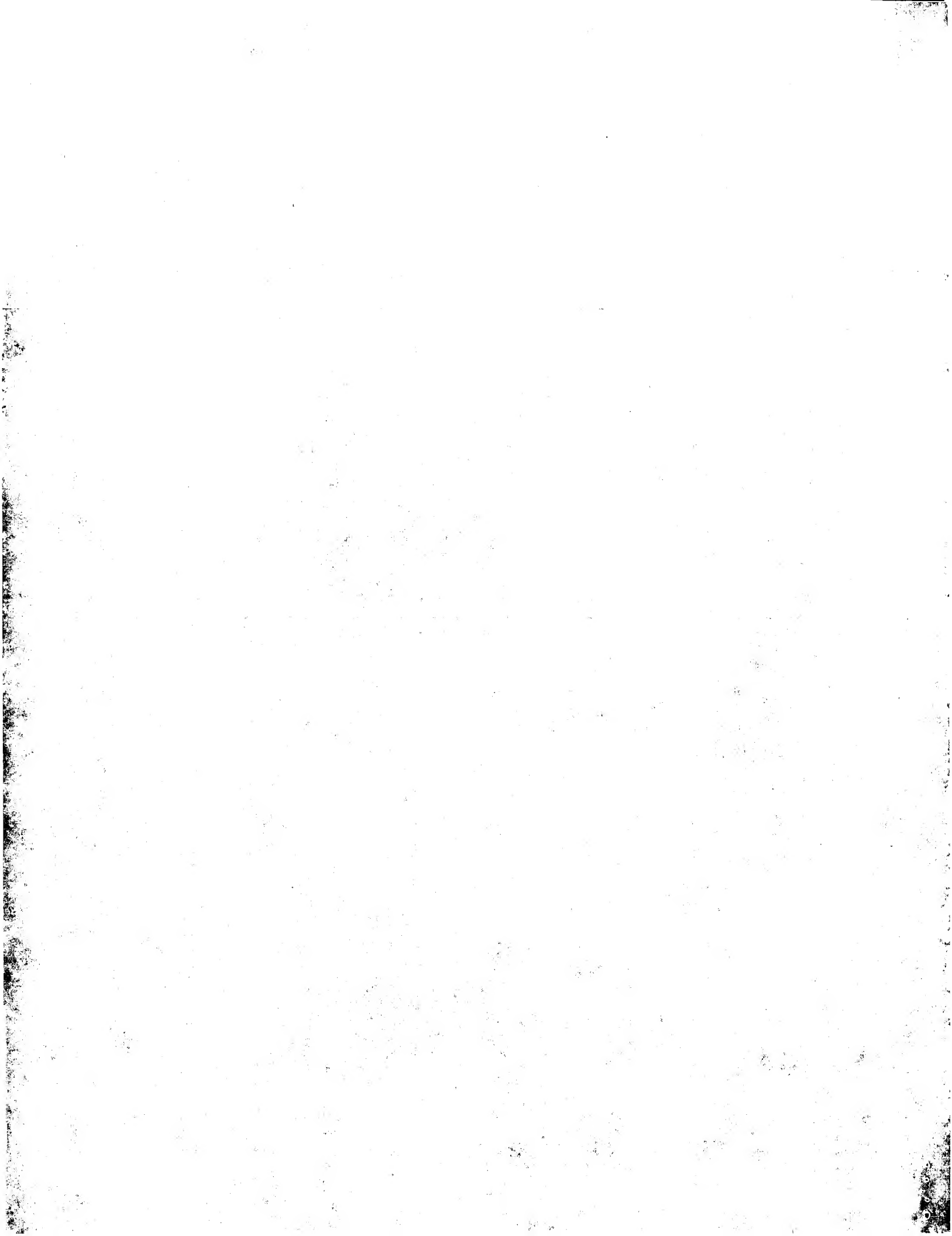


```
DR MEROPS; A01.001; -
DR InterPro; IPR001969; Asp.protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp.1
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59
FT CHAIN 60 387
FT MOD_RES 129 129
FT ACT_SITE 93 93
FT ACT_SITE 276 276
FT DISULFID 106 111
FT DISULFID 267 271
FT DISULFID 310 343
SQ SEQUENCE 387 AA; 42052 MW; 21ADD07782A89585 CRC64;

Query Match 12.2%; Score 295; DB 1; Length 387;
Best Local Similarity 26.1%; Pred.No.5.2e-17;
Matches 97; Conservative 65; Mismatches 123; Indels 86; Gaps 14;

QY 30 YVEMTVGSPQTNLILVDGSSNFAYGAAPHF-----LHRYQQLSSTYRDLRKG 82
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
75 YFGTISIGTPQDFTVIFDTGSSNLWV---PSTYCSSLACALHKRNPEDSSTYQGTSET 131
QY 83 VVYPYTGKWEGLGTDLVSIHPGPNVTVRANIAAITESDKFF-----INGSNWE 132
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
132 LSITYGTGSMTGILGYDTV-----KVGSIEDTNQIFGLSKTEPGLTFLFAPFD 179
QY 133 GILGLAYAEIARPDSDLPPFDLSLVKQTHV~PNLFLQLCGAGFPLNQSEVLASVGGSMI 191
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
180 GILGLAYPSISSDAT~PVFDNNWNEGLYSQDLFSVLSDD-----EKGLVM 227
QY 192 IGGIDHSLYTGSLWYTPIRREWEYEVIVRVEINGQDLKM~DCKEYNDKSIYVDSGTIN 249
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
228 FGGIDSSVYTGSLNWPVSYEGYQWITMDSVSIINGETIACADSC-----QAIVDGTGTS 281
QY 250 LRLPKKYFEAAVKSIKAASSTEKEFPDGFGEQLV~CWQAGTTWNIFPVLISLYLMEV 308
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
282 LTGP-----TSAISNIQSYIGASK-----NLGENVISCASIDSLPDIVF----- 321
QY 309 NQSPRIILPQQYLRPVEDVATSDDCYKFAISQSSTGT-----VMGAVIMEGFYVVD 362
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
322 -----TINGIQYPLPASAYILKEDDDCTSGLEGMNVDTYTGELWILGDVFIQYFTVVD 375
QY 363 RAKRKIGFAVS 373
Db : : : : :
376 RANNQLGLAAA 386
```

Search completed: August 7, 2002, 09:15:11  
Job time: 95 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 50.29 Seconds  
(without alignments)  
1568.616 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEPEPEGRGSGFVEMVD.....CLRLRQCHDDFADDISLLK 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length	ID	Description
1	2419	100.0	532	4 Q9ULS1	Q9ULS1 homo sapien
2	2264.5	93.6	476	4 Q9BYC1	Q9BYC1 homo sapien
3	2161	89.3	457	4 Q9BYC0	Q9BYC0 homo sapien
4	2016.5	83.4	432	4 Q9BYB9	Q9BYB9 homo sapien
5	1407	58.2	266	11 Q9C0U5	Q9C0U5 mus musculus
6	1160.5	48.0	439	4 Q9H2V8	Q9H2V8 homo sapien
7	1146.5	47.4	514	11 Q9UL18	Q9UL18 mus musculus
8	972.5	40.2	468	4 Q9NZL2	Q9NZL2 homo sapien
9	967.5	40.0	396	4 Q9NZL1	Q9NZL1 homo sapien
10	712.5	29.5	213	4 Q9P0D2	Q9P0D2 homo sapien
11	596.5	24.7	255	11 Q9R1P7	Q9R1P7 mus musculus
12	332.5	13.7	391	5 Q9VKP6	Q9VKP6 drosophila
13	332	13.7	354	5 Q9GYX7	Q9GYX7 boophilus m
14	312.5	12.9	386	6 Q9BGU5	Q9BGU5 bos taurus
15	308	12.7	384	13 Q9DEC2	Q9DEC2 xenopus lae
16	308	12.7	385	13 Q9DEC4	Q9DEC4 rana catesb

17	305	12.6	386	6 Q9GMV7	Q9gm7 rhinolophus
18	305	12.6	387	6 Q9GMV8	Q9gmy8 sorex ungui
19	304.5	12.6	372	5 Q9VLK3	Q9vlk3 drosophila
20	304.5	12.6	387	13 Q9DDV5	Q9ddv5 salvelinus
21	304	12.6	387	6 Q9GMV9	Q9gm9 suncus murl
22	303.5	12.5	383	13 Q9DE45	Q9de45 salvelinus
23	303	12.5	383	13 Q9DEC3	Q9dec3 xenopus lae
24	302.5	12.5	376	13 Q9PUR8	Q9pur8 pseudopleur
25	301	12.4	382	13 Q9PRG9	Q9prg9 gallus gall
26	301	12.4	423	5 Q9VKP7	Q9vkp7 drosophila
27	298.5	12.3	384	13 Q9I322	Q9i322 rana catesb
28	295.5	12.2	386	6 Q9GMV6	Q9gm6 canis famil
29	294	12.2	396	13 Q93428	Q93428 chionodraco
30	290.5	12.0	381	6 Q9GK11	Q9gk11 camelus dro
31	288	11.9	399	13 Q93458	Q93458 podarcis si
32	287.5	11.9	444	5 Q21966	Q21966 caenorhabdi
33	284	11.7	398	13 P87370	P87370 oncorhynch
34	284	11.7	427	5 P91802	P91802 schistosoma
35	281	11.6	378	13 Q9FUR9	Q9fur9 pseudopleur
36	279.5	11.6	390	6 Q9GK10	Q9gk10 camelus dro
37	278	11.5	370	6 Q9TTW1	Q9ttw1 bos taurus
38	278	11.5	399	13 Q9DD89	Q9dd89 brachydanio
39	278	11.5	422	5 Q96906	Q96906 onchocerca
40	277.5	11.5	380	6 Q28950	Q28950 sus scrofa
41	277	11.5	446	5 Q9N9H3	Q9n9h3 necator ane
42	276	11.4	389	13 Q9PWK1	Q9pwk1 gallus gall
43	276	11.4	389	13 Q9W643	Q9w643 gallus gall
44	276	11.4	396	13 Q9DEX3	Q9dex3 clipea hare
45	275	11.4	385	6 Q29080	Q29080 sus scrofa

## ALIGNMENTS

RESULT 1

Q9ULS1 PRELIMINARY; PRT; 532 AA.

ID Q9ULS1; 01-MAY-2000 (TRENBLrel. 13, Created)

AC Q9ULS1; 01-OCT-2001 (TRENBLrel. 18, Last sequence update)

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)

DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

DE KIAA1149 PROTEIN (FRAGMENT).

GN KIAA1149.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=20039618; PubMed=10574461;

RA Hirosewa M., Nagase T., Ishikawa K., Kikuno R., Ohara O.;

RT "Characterization of cDNA clones selected by the GeneMark analysis

RL from size-fractionated cDNA libraries from human brain.";

DR EMBL; AB032975; BAA86463.2; --

DR HSSP; P56272; IAM5.

DR MEROPS; A01.004; --

DR InterPro; IPR001969; Asp\_protease.

DR InterPro; IPR001461; Pepsin.

DR Pfam; PF00026; asp; 1.

DR PRINTS; PRO0792; PEPSTN.

DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.

FT NON\_TER 1

SQ SEQUENCE 532 AA; 58720 MW; 98B135D0D5FBD2E8 CRC64;

Query Match 100.0%; Score 2419; DB 4; Length 532;  
Best Local Similarity 100.0%; Pred. No. 7.3e-198;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ETDEPEPEGRGSGFVEMVDNLCKSQGYVEMTVGSPPTNLILVDTGSSNFVGAAP 60  
|||||

Db 106 HPFLHRYQRLSSTYRDLRKGVVVPYTOCKWEGELGTLVSIPIHGPNTVTRANIAAITE 165  
Qy 121 SDKFFINGSNWEGILGLAYAEIARPDLSLPPFDLSLVKQTHVPNLFSLQCGAGFPLNQS 180  
Db 166 SDKFFINGSNWEGILGLAYAEIAR-----LCGAGFPLNQS 200  
Qy 181 EVLASVGSMTIIGDHSLYTSGSLWYTPIRREWYVEYIIIVRVEINGQDLKMDCKEYNDK 240  
Db 201 EVLASVGSMTIIGDHSLYTSGSLWYTPIRREWYVEYIIIVRVEINGQDLKMDCKEYNDK 260  
Qy 241 SIVDSGTTNLRPLKPKVFAAASKAASSTKFPDGFGLVGEOLVCMQAGTTPWNIFFVIS 300  
Db 261 SIVDSGTTNLRPLKPKVFAAASKAASSTKFPDGFGLVGEOLVCMQAGTTPWNIFFVIS 320  
Qy 301 LYLMEVTNQSFRTILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVYV 360  
Db 321 LYLMEVTNQSFRTILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVYV 380  
Qy 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 420  
Db 381 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 440  
Qy 421 CALFMLPLCLMVCMQRCRLRCLRQHQHDDFADDISLLK 456  
Db 441 CALFMLPLCLMVCMQRCRLRCLRQHQHDDFADDISLLK 476

RESULT 3  
Q9BYC0 PRELIMINARY; PRT; 457 AA.  
AC Q9BYC0; (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 19, Last annotation update)  
DE BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME TYPE C).  
DE BACE.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabira T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP cleaving enzyme (BACE) and their effect on amyloid beta-peptide production.";  
RL Neurosci. Lett. 307:9-12(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EXOCRINE PANCREAS;  
RA Zaccchetti D., De Pietri Tonelli D., Schnurbus R.;  
RT "New beta-site APP cleaving enzyme isoform (BACE-1C) obtained from human pancreas.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB050437; BAB40932.1; -;  
DR EMBL; AF338817; AAK38375.1; -;  
DR HSSP; P32329; 1YPS  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN.1.  
SQ SEQUENCE 457 AA; 51068 MW; C794C9A9E85FE7A2 CRC64;

Query Match 89.3%; Score 2161; DB 4; Length 457;  
Best Local Similarity 90.4%; Pred. No. 6.1e-176;  
Matches 412; Conservative 0; Mismatches 0; Indels 44; Gaps 1;  
Qy 1 ETDEEPEEPCRRGSGFVEMVDNLGRKSGQYVYVMTVSGPQTLNILDVTGSSNFVGAAP 60

Db 77 ETDEEPEEPCRRGSGFVEMVDNLGRKSGQYVYVMTVSGPQTLNILDVTGSSNFVGAAP 136  
Qy 61 HPFLHRYQRLSSTYRDLRKGVVVPYTOCKWEGELGTLVSIPIHGPNTVTRANIAAITE 120  
Db 137 HPFLHRYQRLSSTYRDLRKGVVVPYTOCKWEGELGTLVSIPIHGPNTVTRANIAAITE 196  
Qy 121 SDKFFINGSNWEGILGLAYAEIARPDLSLPPFDLSLVKQTHVPNLFSLQCGAGFPLNQS 180  
Db 197 SDKFFINGSNWEGILGLAYAEIARPDLSLPPFDLSLVKQTHVPNLFSLQCGAGFPLNQS 256  
Qy 181 EVLASVGSMTIIGDHSLYTSGSLWYTPIRREWYVEYIIIVRVEINGQDLKMDCKEYNDK 240  
Db 257 EVLASVGSMTIIGDHSLYTSGSLWYTPIRREWYVEYIIIVRVEINGQDLKMDCKEYNDK 316  
Qy 241 SIVDSGTTNLRPLKPKVFAAASKAASSTKFPDGFGLVGEOLVCMQAGTTPWNIFFVIS 300  
Db 317 SIVDSGTTNLRPLKPKVFAAASKAASSTKFPDGFGLVGEOLVCMQAGTTPWNIFFVIS 376  
Qy 301 LYLMEVTNQSFRTILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVYV 360  
Db 377 LYLMEVTNQSFRTILPQQYLRPVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGYVYV 436  
Qy 361 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 420  
Db 437 FDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDDESTLMTIAYVMAAI 496  
Qy 421 CALFMLPLCLMVCMQRCRLRCLRQHQHDDFADDISLLK 456  
Db 497 CALFMLPLCLMVCMQRCRLRCLRQHQHDDFADDISLLK 532

RESULT 2  
Q9BYC1 PRELIMINARY; PRT; 476 AA.  
AC Q9BYC1; (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 19, Last annotation update)  
DE BETA-SITE APP CLEAVING ENZYME I-476.  
DE BACE.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RX MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabira T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP cleaving enzyme (BACE) and their effect on amyloid beta-peptide production.";  
RL Neurosci. Lett. 307:9-12(2001).  
DR EMBL; AB050436; BAB40931.1; -;  
DR HSSP; P32329; 1YPS.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN.1.  
SQ SEQUENCE 476 AA; 52907 MW; 6C8C87F8A953AF66 CRC64;

Query Match 93.6%; Score 2264.5; DB 4; Length 476;  
Best Local Similarity 94.5%; Pred. No. 9.5e-185;  
Matches 431; Conservative 0; Mismatches 0; Indels 25; Gaps 1;  
Qy 1 ETDEEPEEPCRRGSGFVEMVDNLGRKSGQYVYVMTVSGPQTLNILDVTGSSNFVGAAP 60  
Db 46 ETDEEPEEPCRRGSGFVEMVDNLGRKSGQYVYVMTVSGPQTLNILDVTGSSNFVGAAP 105  
Qy 61 HPFLHRYQRLSSTYRDLRKGVVVPYTOCKWEGELGTLVSIPIHGPNTVTRANIAAITE 120

```
Db 46 ETDEEPEPCRRGGSFVEMVDNLKSGSQGYVYVMTGSPQTLNLLVDGSSNFVGAAP 105
QY 61 HPFLHRYQRLSSYTRDLRKGYYVYPTQKGWEGELGTDLVSIPIHGPNVTVRANIAAITE 120
Db 106 HPFLHRYQRLSSYTRDLRKGYYVYPTQKGWEGELGTDL----- 145
QY 121 SKFFINGSNWEGILGLAYAEIARPDSDLPPFDSLVKQTHVNPFLSFLQLCGAGPPLNQS 180
Db 146 -----PDSLEPFDSLVKQTHVNPFLSFLQLCGAGPPLNQS 181
QY 181 EYLASVGGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240
Db 182 EYLASVGGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 241
QY 241 SIVDSGTTNLRPLPKVFEAAVKSIAAASSTKFPDGFGLWGEQVLCWQAGTTPWNIFPVIS 300
Db 242 SIVDSGTTNLRPLPKVFEAAVKSIAAASSTKFPDGFGLWGEQVLCWQAGTTPWNIFPVIS 301
QY 301 LYLMGEVNTQSFRTITLPPQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 302 LYLMGEVNTQSFRTITLPPQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 361
QY 361 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 420
Db 362 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 421
QY 421 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 456
Db 422 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 457

RESULT 4
Q9BYB9 ID Q9BYB9 PRELIMINARY; PRT; 432 AA.
AC Q9BYB9
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-432.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21408467; Pubmed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production.";
RL Neurosci. Lett. 307:9-12(2001).
RL EMBL; AB050438; BAB40933.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSTN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 432 AA; 48212 MW; 96FC81B6F0ED01B CRC64;

Query Match 83.4%; Score 2016.5; DB 4; Length 432;
Best Local Similarity 84.9%; Pred. No. 1.2e-163;
Matches 387; Conservative 0; Mismatches 0; Indels 69; Gaps 1;

QY 1 ETDEEPEPCRRGGSFVEMVDNLKSGSQGYVYVMTGSPQTLNLLVDGSSNFVGAAP 60
Db 46 ETDEEPEPCRRGGSFVEMVDNLKSGSQGYVYVMTGSPQTLNLLVDGSSNFVGAAP 105
QY 61 HPFLHRYQRLSSYTRDLRKGYYVYPTQKGWEGELGTDLVSIPIHGPNVTVRANIAAITE 120
```

```
Db 106 HPFLHRYQRLSSYTRDLRKGYYVYPTQKGWEGELGTDL----- 145
QY 121 SDKFFINGSNWEGILGLAYAEIARPDSDLPPFDSLVKQTHVNPFLSFLQLCGAGPPLNQS 180
Db 146 -----LCGAGPPLNQS 156
QY 181 EYLASVGGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 240
Db 157 EYLASVGGSMIIGGIDHSLYTSGLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDK 216
QY 241 SIVDSGTTNLRPLPKVFEAAVKSIAAASSTKFPDGFGLWGEQVLCWQAGTTPWNIFPVIS 300
Db 217 SIVDSGTTNLRPLPKVFEAAVKSIAAASSTKFPDGFGLWGEQVLCWQAGTTPWNIFPVIS 276
QY 301 LYLMGEVNTQSFRTITLPPQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 277 LYLMGEVNTQSFRTITLPPQYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 336
QY 361 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 420
Db 337 FDRARRKIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDSTLMTIAYVMAAI 396
QY 421 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 456
Db 397 CALFMLPLCLMVCQWRCRLRQHQHDDFADDISLLK 432

RESULT 5
Q9CUU5 ID Q9CUU5 PRELIMINARY; PRT; 266 AA.
AC Q9CUU5
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME (FRAGMENT).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=BRAIN;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014390; BAB29317.1; -.
DR MEROPS; A01.004; -.
DR MGD; MGI:1346542; Bace.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 2.
FT NON_TER 1
SQ SEQUENCE 266 AA; 30201 MW; B913FDA8ADAB4238 CRC64;
```

	Query Match	58.2%; Score 1407; DB 11; Length 266;	
	Best Local Similarity	98.9%; Pred. No. 5.9e-112;	
	Matches 263; Conservative	1; Mismatches 2; Indels 0; Gaps 0;	
Qy	191	ITGGIDHSLYTGLWYTPIRREWWYEVIIVRVEINGODLKMDCKEYNVDKSIDVSGTTNL 250	
Dd	1	ITGGIDHSLYTGLWYTPIRREWWYEVIIVRVEINGODLKMDCKEYNVDKSIDVSGTTNL 60	
Qy	251	RLPKKVFEEAAVKSIKAASSTERFPDGFVLGEOLVCWQAGTTPWNIPFVVISLVMGEVTNQ 310	
Dd	61	RLPKKVFEEAAVKSIKAASSTERFPDGFVLGEOLVCWQAGTTPWNIPFVVISLVMGEVTNQ 120	
Qy	311	SFRITILPQOYLRLPVEDVATSDODCYKFAISOSSTGTVMGAVIMEGYVVFDRARKRIGF 370	
Dd	121	SFRITILPQOYLRLPVEDVATSDODCYKFAVSQSSTGTVMGAVIMEGYVVFDRARKRIGF 180	
Qy	371	AVSACHVDEFTAAVEGPFVTLDMEDCCYNIPTQDDESTLMTIAYVMAAICALEFMLPLCL 430	
Dd	181	AVSACHVDEFTAAVEGPFVTADEMDCGNIPQTDESTLMTIAYVMAAICALFMLPLCL 240	
Qy	431	MVCWRCLRCLROQHDDDFADDISLLK 456	
Dd	241	MVCWRCLRCLRHQHDDDFADDISLLK 266	
	RESULT	6	
ID	Q9H2V8	PRELIMINARY; PRT; 439 AA.	
AC	Q9H2V8;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	CDAL3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PHEOCHROMOCYTOMA;		
RA	Li Y., Huang Q., Peng, Y, Song H., Yu Y., Xu S., Ren S., Chen Z.,		
RA	Han Z.;		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.		
RL	EMBL: AF212252; AAC41783.1; --		
DR	HSP: P00797; 2REN		
DR	InterPro: IPR001969; Asp_protease.		
DR	InterPro: IPR001461; Pepsin.		
DR	Pfam: PF00026; asp; 3.		
DR	PRINTS; PR00792; PEPSIN.		
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.		
SQ	SEQUENCE 439 AA; 48275 MW; 02EC0E0E50F11602 CRC64;		
	Query Match	48.08; Score 1160.5; DB 40; Length 439;	
	Best Local Similarity	50.1%; Pred. No. 1.3e-90;	
	Matches 220; Conservative	78; Mismatches 134; Indels 7; Gaps 4;	
Qy	18	MVDNLRGKSGGYVEMTVGSPPTNLILVDTGSSNFVAGAAPHPFLHRYQRQLSSTYR 77	
Dd	1	MVDNLQDGSGRGYLEMLICTPPKQLIQLVDTGSSNFVAGAGTHSYDIYDTERSSYR 60	
Qy	78	DLRGVYPVYTGKWEGELGTDLSIPHGPNNVTVRANIAATESDKPFINGSNWEGILGL 137	
Dd	61	SKGFDVTVKYTGQGWTFGVGEDLVTPKGFTSLFNLTATIFESENFPLPGIKWNGILGL 120	
Qy	138	AAETARPDDSLPEFFDSLVKOTHPNLFSLQLCGAGFPLNQSESVLASVGSMIIIGIDH 197	
Dd	121	AYATLAKPSSLETFEFDLSVTQANIPIVNSWMQMGAGLPVAGS---GTNGSLVLGGIEP 177	
Qy	198	SLYTGLSWYTPIRREWWYEVIIVRVEINGODLKMDCKEYNVDKSIDVSGTTNLRLPKKVF 257	
Dd	178	SLYKGDGIWTPIKEWYIQIELKLEIGQSNLDRCYNADKAIVDSGTTLLRPKPQKVF 237	



```
ID Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=CORD BLOOD;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang O.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1;
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR NON_TER 1
FT SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;
SQ

Query Match 29.5%; Score 712.5; DB 4; Length 213;
Best Local Similarity 83.5%; Pred. No. 7.8e-53;
Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

Qy 193 GGIDHSLYTSGSLWTPTRREHYEIVIRVEINGQDLKMDCKEYNDKSIYDVSQTTNRL 252
Db 1 GGIDHSLYTSGSLWTPTRREHYEIVIRVEINGQDLKMDCKEYNDKSIYDVSQTTNRL 60
Qy 253 PKKFEAAVKSIAASSTKEKPPDGFGLGEQLVCWQAGTTPWNIIPVLSLYLMGEVTVNOSF 312
Db 61 PKKFEAAVKSIAASSTKEKPPDGFGLGEQLVCWQAGTTPWNIIPVLSLYLMGEVTVNOSF 120
Qy 313 RITILPOQLRPVEDVATSDQDCKYFAISQSSTGTVMGAVIMEG 356
Db 121 RITILPOQLRP-----WKMMPPRKTVTVCHLTVIHG 153

RESULT 11
Q9R1P7 PRELIMINARY; PRT; 255 AA.
AC Q9R1P7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC Acgario M., Fumagalli P., Taramelli R., Ottolenghi S.;
RA "Cloning of a gene from chromosome 21 Down Region encoding a potential
RA transmembrane protease.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF051150; AAD45964.1;
DR MEROPS; A01.041;
DR InterPro; IPR001969; Asp-protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 2.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW Protease.
FT NON_TER 1
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;

Query Match 24.7%; Score 596.5; DB 11; Length 255;
Best Local Similarity 47.8%; Pred. No. 8e-43;
```

```
Matches 109; Conservative 44; Mismatches 74; Indels 1; Gaps 1;

Qy 205 WYTPIRREWYEVIVRVEINGQDLKMDCKEYNDKSIYDVSQTTNRLPKKVFEEAAVSI 264
Db 1 WYTPIRREWYEVIVRVEINGQDLKMDCKEYNDKSIYDVSQTTNRLPKKVFEEAAVSI 60
Qy 265 KRAASSTKEKPPDGFGLGEQLVCWQAGTTPWNIIPVLSLYLMGEVTVNOSFRITILPOQLRP 324
Db 61 ARTSLPEESDGFGLGEQLVCWQAGTTPWNIIPVLSLYLMGEVTVNOSFRITILPOQLRP 120
Qy 325 VEDVATSDQDCKYFAISQSSTGTVMGAVIMEGTVVDFRRKRIGFAVSACHVHDEFRTA 384
Db 121 MMGAGENY-ECYRGFSSSTNALVIGATVMEGTVVDFRRKRIGFAVSACHVHDEFRTA 179
Qy 385 AVBGPFTVLMDCGYNIPQDTEMTLTAYVMAATCALFMLEPLCLMV 432
Db 180 EISGPFSTEDIASNCVPAQALNEPILWIVSYALMSVCGAILLVILL 227

RESULT 12
Q9VKP6 PRELIMINARY; PRT; 391 AA.
AC Q9VKP6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG17134 PROTEIN.
GN CG17134.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M., Reese M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```





Query Match 12.9%; Score 312.5; DB 6; Length 386;  
Best Local Similarity 28.0%; Pred. No. 2.5e-18;  
Matches 113; Conservative 69; Mismatches 131; Indels 91; Gaps 20;

QY 8 EPG--RRGSFVEMVDNLRKSGQGYVEMTGVSPPTNLIVDTGSSNFAVGAAPHPFL-- 64  
DB 35 EPAVRQGPPELLKNYMDAQ---YIGIGIGTTPQCTVTVFDGNSANLWVPSIHKLLDI 91  
QY 65 ----HRYQRLSTYRDLRGVY--VPYTGKWEGLGTLVSIPIHGPN-----VTVR 112  
DB 92 ACWTHRKYNSDKSTY--VKNGTTFDHYGSGSLGSLQSDTVSPVPCNPSSSPGGVTQ 149  
QY 113 ANI--AAITESDREFFINGSNWEGILGAXAFIARPDLSLEPFDLSLVKQTHV--PNLFSLQ 169  
DB 150 RQTFGEAIKQPGVVFI--AAKFGGLGWAYPRIS--VNNVLPVFDNLQCKLVKKNFVS-- 204  
QY 170 LCGAGFLPNQSEVLASVGGSMIGGIDHSLYTGSMTPIRREWYEVIIVRVEINGQDL 229  
DB 205 -----FFLNR-DPKAOPGGELMLGSTDYKYYRGSLSMFHNTRQAYWQIHMDQLDV-GSSL 257  
QY 230 KMDCKEYNDKSIDVSTTNLRPLPKVFEAAVKSIIKAASSTEKEPDPGFWLGEQLVCHQAG 289  
DB 258 TV-CK--GGCEAIVDTGTSILVGPVEEVREIQKAI-----G 290  
QY 290 TTPWNIPFVISLXLMGEVTNQSFRITILPOQYLR-PVEDVATSDDCYKFAISOSSTGT- 347  
DB 291 AVP-----LIQGYMIPCEKVSLSLPQVTVKLGKDYAXSPED-YALKVSAQGTVC 340  
QY 348 -----VMGAVIMEGFYVVFDRARKRIGFAVSA 374  
DB 341 LSGPMGMDIPPPGGPLMLGLDVFIGRYTTFDRDQNRVGLAEAA 384

RESULT 15  
Q9DEC2 Q9DEC2 PRELIMINARY; PRT; 384 AA.  
AC Q9DEC2;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PEPSINOGEN A.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasu S.;  
RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog  
RT Rana catesbeiana."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR EMBL; AB045380; BAB20798.1; -.  
DR HSP; P00790; 1PSN.  
DR MEROPS; A01.001; -.  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
KW Aspartyl protease; Hydrolase.  
SQ SEQUENCE 384 AA; 41888 MW; 729DD2E7953D9072 CRC64;

Query Match 12.7%; Score 308; DB 13; Length 384;  
Best Local Similarity 27.5%; Pred. No. 6e-18;  
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;

QY 30 YVEMTVCSPPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSTYRDLRKG 82  
DB 35 EPAVRQGPPELLKNYMDAQ---YIGIGIGTTPQCTVTVFDGNSANLWVPSIHKLLDI 91

DB 72 YYGTSIGTPPQETVILFDGTSANLWV---PSVYCSSQACSNHNRNFPQOSSTFQATNTP 128  
QY 83 VYYPYTOGKWEGLGTLVSIPIHGPVTVRANIAAITESDK-FFINGSNWEGILGAYAE 141  
DB 129 VSIQYGTGSMGSGFLGYDTLQV---GNIQISNQMEGLSESEPGSELYYSPFDGILGLAFPS 185  
QY 142 IARPDLSLEPFDLSLVKQTHV-PNLSLQLCGAGFLPNQSEVLASVGGSMIGGIDHSLY 200  
DB 186 IA--SSQATPVFDNMWSQGLIPQNLFSVYLSDDG-----QTGSYVLFGGVDSYY 233  
QY 201 TGSWYTPIRREWYEVIIVRVEINGQDL--KMDCKEYNDKSIDVSGTTLNLRPLPKVFE 258  
DB 234 SGLSNWVPLTAETWQITLDSVINGQVIACQSC-----QAIVDTGTSMTGPTPI- 286  
QY 259 AAVKSIIKAASSTEKEPDPGFWLGEOLV-CWQAGTTPWNIPFVISLXLMGEVTNQSFRITIL 317  
DB 287 ANIQNYIGASODSN-----GQIVINCNNISNPTIVF-----TIN 321  
QY 318 PQQY-LRPVEDVATSDDCYK-FAISOSSTGT-----VMGAVIMEGFYVVFDRARKRIGFA 371  
DB 322 GVQYPLSPSAVVRQNGQCGSSGFCQAMNLPNTNSGDLWILGDVFIROYFTVDFRANNYYAIA 381

Search completed: August 7, 2002, 09:17:12  
Job time: 216 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:35 ; Search time 22.58 Seconds  
(without alignments)  
493.272 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEPEEPGRGSGFVEMVD.....CLRLRQHQHDDFADDISLLK 456

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2413	99.8	501	4	US-09-009-191-2
2	2320	95.9	774	4	US-09-009-191-4
3	1171.5	48.4	518	3	US-08-999-723-2
4	1171.5	48.4	518	4	US-09-434-427-2
5	1146.5	47.4	514	4	US-09-717-432-2
6	1146.5	47.4	514	4	US-09-912-484-2
7	299.5	12.4	396	1	US-08-208-007A-13
8	299.5	12.4	396	1	US-09-032-523-9
9	295.5	12.2	412	1	US-08-208-007A-12
10	285.5	12.2	412	4	US-08-974-691-4
11	279.5	11.6	458	6	5217891-15
12	273.5	11.3	409	1	US-09-640-305-6
13	273.5	11.3	409	1	US-08-360-673-6
14	273.5	11.3	427	2	US-08-846-021A-8
15	270	11.2	410	1	US-08-088-633-2
16	270	11.2	410	1	US-08-245-756-2
17	270	11.2	410	1	US-08-441-750-2
18	270	11.2	410	1	US-08-441-751-2
19	270	11.2	410	5	PCT-US92-02521-2
20	250	10.3	349	4	US-09-032-523-3
21	227	9.4	398	1	US-08-328-314-2
22	227	9.4	398	1	US-08-731-045-2
23	212	8.6	437	3	US-08-974-691-3
24	210	8.7	397	4	US-09-079-415-2
25	196.5	8.1	430	1	US-08-535-237-2
26	195.5	8.1	427	1	US-07-958-222A-2
27	193	8.0	330	3	US-08-115-753-1

28	193	8.0	419	3	US-08-115-753-2	Sequence 2, Appli
29	193	8.0	419	3	US-08-115-753-3	Sequence 33, Appli
30	187	7.7	420	4	US-09-008-271A-4	Sequence 4, Appli
31	187	7.7	420	4	US-08-974-691-8	Sequence 8, Appli
32	184.5	7.6	395	1	US-08-723-938-3	Sequence 3, Appli
33	184.5	7.6	395	2	US-09-080-538-3	Sequence 3, Appli
34	184	7.6	445	4	US-08-974-691-6	Sequence 6, Appli
35	184	7.6	445	4	US-08-974-691-2	Sequence 6, Appli
36	149	6.2	437	4	US-09-353-332-2	Sequence 2, Appli
37	129.5	5.4	140	3	US-09-211-631-13	Sequence 2, Appli
38	129.5	5.4	140	4	US-09-265-628-13	Sequence 13, Appli
39	129.5	5.4	140	4	US-09-001-141-11	Sequence 11, Appli
40	129.5	5.4	140	4	US-09-532-803-6	Sequence 6, Appli
41	129.5	5.4	140	4	US-09-653-403-14	Sequence 14, Appli
42	97	4.0	1030	4	US-09-091-117-2	Sequence 2, Appli
43	95.5	3.9	280	4	US-09-160-246-14	Sequence 14, Appli
44	92.5	3.8	1097	2	US-08-680-326-39	Sequence 39, Appli
45	88	3.6	746	2	US-08-838-219B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-009-191-2  
; Sequence 2, Application US/09009191  
; Patent No. 6319689  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID  
; APPLICANT: CHAPMAN, CONRAD  
; APPLICANT: MURPHY, KAY  
; APPLICANT: SMITH, TRUDI  
; TITLE OF INVENTION: ASP2  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,191  
; FILING DATE: 20-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UK 9701684.4  
; FILING DATE: 28-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23,031  
; REFERENCE/DOCKET NUMBER: GH-70368  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 501 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-009-191-2

Query Match 99.8%; Score 2413; DB 4; Length 501;  
Best Local Similarity 99.8%; Pred. No. 3.3e-251;



QY 9 PGRGSEFVEMVNDLRKSGQYVYVEMTGGSPQTLNILDVTGSSNFVAGAAPHLHRY 68  
Db 71 PACAAFLAMVDNLQSDSGRYYLEMLIGTPQKQLIYLDVTGSSNFVAGTPHSYIDTYF 130  
QY 69 QRLSSTYRDLRKGVYVPTQGWKEGELGDLVSIHPGNVTVRANIAAITESDKFFING 128  
Db 131 DTERSTYRSKGFDTVKYTGTSWTFGVGEDLVTPKGFNTSFLVNIATIPESNFELPG 190  
QY 129 SNWEGILGLAYAEIARPDLSLPPFDDSLVKQTHVNLFSQLCGAGFPLNQSEVLASVG 188  
Db 191 IKWNGILGLAYATLAKPSSLETFDLSVTQANTIPNVFSMQMCGAGLPVAGS---GTNG 247  
QY 189 SMIIIGDHSLSYGLWYTPIRREYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSGTT 248  
Db 248 SLVLGGIEPSLYKGIWYTPKEEYQIETLKEIGQSLNLDREYNADKAIVDSGTT 307  
QY 249 NLRLPKKVFEEAAVKSIAASSTKPEKPDGFWLGEQLVCMQAGTTPWNIFFPVISLYLMGEVT 308  
Db 308 LLRLPKQKVDVAVAEARASLIPEFSDGFWTGSQACWTNSETPWSYFPKISYIYLRDENS 367  
QY 309 NOSFRITILPOQLRPVEDVATSDQDCKYFAISQSSSTGTVMGAVIMEGFYVYVDFRARKRI 368  
Db 368 SRSFRITILPOLYIQPMGAGLNY-ECYRFGISPTNALVIGATVMGEGFYVDFRARKRV 426  
QY 369 GFAYSACHVHDEFRTAAVEGFPVTLDMEDCGYNIPQTDDESTLMTIAYVMAAIC-ALFMLP 427  
Db 427 GFASPCAEIAGAAVSEISGPFSTEDVASNCVPAQSLSEPLIWIYVYALMSVCGAILLVL 486  
QY 428 LCLMVCWRCRLCRLRQHQHDDFADDSILL 455  
Db 487 IVLLLLPFCQR--RPRDPEVNVDESSL 512

## RESULT 4

US-09-434-427-2  
; Sequence 2, Application US/09434427  
; Patent No. 6162630  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID J.  
; APPLICANT: SOUTHAN, CHRISTOPHER  
; APPLICANT: CHAPMAN, CONRAD G.  
; APPLICANT: EVANS, JOANNE R.  
; TITLE OF INVENTION: ASP1  
; FILE REFERENCE: GH-70262-D1  
; CURRENT APPLICATION NUMBER: US/09/434,427  
; CURRENT FILING DATE: 1999-11-04  
; EARLIER FILING DATE: 1997-10-06  
; EARLIER APPLICATION NUMBER: UK 9626022.9  
; EARLIER FILING DATE: 1996-12-14  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 518  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-434-427-2

Query Match 48.4%; Score 1171.5; DB 4; Length 518;  
Best Local Similarity 49.6%; Pred. No. 2.2e-117;  
Matches 222; Conservative 80; Mismatches 139; Indels 7; Gaps 4;

QY 9 PGRGSEFVEMVNDLRKSGQYVYVEMTGGSPQTLNILDVTGSSNFVAGAAPHLHRY 68  
Db 71 PACAAFLAMVDNLQSDSGRYYLEMLIGTPQKQLIYLDVTGSSNFVAGTPHSYIDTYF 130  
QY 69 QRLSSTYRDLRKGVYVPTQGWKEGELGDLVSIHPGNVTVRANIAAITESDKFFING 128  
Db 131 DTERSTYRSKGFDTVKYTGTSWTFGVGEDLVTPKGFNTSFLVNIATIPESNFELPG 190  
QY 129 SNWEGILGLAYAEIARPDLSLPPFDDSLVKQTHVNLFSQLCGAGFPLNQSEVLASVG 188  
Db 191 IKWNGILGLAYATLAKPSSLETFDLSVTQANTIPNVFSMQMCGAGLPVAGS---GTNG 247

Db 191 IKWNGILGLAYATLAKPSSLETFDLSVTQANTIPNVFSMQMCGAGLPVAGS---GTNG 247  
QY 189 SMIIIGDHSLSYGLWYTPIRREYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSGTT 248  
Db 248 SLVLGGIEPSLYKGIWYTPKEEYQIETLKEIGQSLNLDREYNADKAIVDSGTT 307  
QY 249 NLRLPKKVFEEAAVKSIAASSTKPEKPDGFWLGEQLVCMQAGTTPWNIFFPVISLYLMGEVT 308  
Db 308 LLRLPKQKVDVAVAEARASLIPEFSDGFWTGSQACWTNSETPWSYFPKISYIYLRDENS 367  
QY 309 NOSFRITILPOQLRPVEDVATSDQDCKYFAISQSSSTGTVMGAVIMEGFYVYVDFRARKRI 368  
Db 368 SRSFRITILPOLYIQPMGAGLNY-ECYRFGISPTNALVIGATVMGEGFYVDFRARKRV 426  
QY 369 GFAYSACHVHDEFRTAAVEGFPVTLDMEDCGYNIPQTDDESTLMTIAYVMAAIC-ALFMLP 427  
Db 427 GFASPCAEIAGAAVSEISGPFSTEDVASNCVPAQSLSEPLIWIYVYALMSVCGAILLVL 486  
QY 428 LCLMVCWRCRLCRLRQHQHDDFADDSILL 455  
Db 487 IVLLLLPFCQR--RPRDPEVNVDESSL 512

## RESULT 5

US-09-717-432-2  
; Sequence 2, Application US/09717432  
; Patent No. 6291223  
; GENERAL INFORMATION:  
; APPLICANT: ZHU, YUAN  
; APPLICANT: LI, XIAOTONG  
; APPLICANT: CHRISTIE, GARY  
; APPLICANT: POWELL, DAVID J.  
; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mAsp1)  
; FILE REFERENCE: GP-70663  
; CURRENT APPLICATION NUMBER: US/09/717,432  
; CURRENT FILING DATE: 2000-11-21  
; PRIOR APPLICATION NUMBER: 60/166,974  
; PRIOR FILING DATE: 1999-11-23  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: MUS MUSCULUS  
US-09-717-432-2

Query Match 47.4%; Score 1146.5; DB 4; Length 514;  
Best Local Similarity 50.7%; Pred. No. 1.1e-114;  
Matches 216; Conservative 75; Mismatches 130; Indels 5; Gaps 3;

QY 8 EPCR-RGSEFVEMVNDLRKSGQYVYVEMTGGSPQTLNILDVTGSSNFVAGAAPHLHRY 66  
Db 65 EPVRATANFLAMVDNLQSDSGRYYLEMLIGTPQKQLIYLDVTGSSNFVAGAPHSYIDT 124  
QY 67 YYQRQLSSTYRDLRKGVYVPTQGWKEGELGDLVSIHPGNVTVRANIAAITESDKFFI 126  
Db 125 YFDESSTYHSKGFDTVKYTGTSWTFGVGEDLVTPKGFNTSFLVNIATIPESNFEL 184  
QY 127 NGSNNEGILGLAYAEIARPDLSLPPFDDSLVKQTHVNLFSQLCGAGFPLNQSEVLASV 186  
Db 185 PGIKWNGILGLAYAEIARPDLSLPPFDDSLVKQTHVNLFSQLCGAGFPLNQSEVLASV 241  
QY 187 GGSMLIIGDHSLSYGLWYTPIRREYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSG 246  
Db 242 GGSIVLGGIEPSLYKGIWYTPKEEYQIETLKEIGQSLNLDREYNADKAIVDSG 301  
QY 247 TTNLRLPKKVFEEAAVKSIAASSTKPEKPDGFWLGEQLVCMQAGTTPWNIFFPVISLYLMGE 306  
Db 302 TTLRLPKQKVDVAVAEARASLIPEFSDGFWTGSQACWTNSETPWSYFPKISYIYLRDE 361  
QY 307 VTNQSFRTILPOQLRPVEDVATSDQDCKYFAISQSSSTGTVMGAVIMEGFYVYVDFRARK 366  
Db 368 SRSFRITILPOLYIQPMGAGLNY-ECYRFGISPTNALVIGATVMGEGFYVDFRARKRV 426

Db 362 NASRSFRITILPOLYTOPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGFVYVDFDRAOR 420  
QY 367 RIGFAVSACHVHDEFTAAVEGFVTLDMEDCGYNIPODESTLMTIAYVMAAICALFML 426  
Db 421 RVGFVSPCAEIGTGVSEISGPFSTEDIASNCVPAQALNEPILMTIVSVYALMSVCGAAILL 480  
QY 427 PLCLMV 432  
Db 481 VLILL 486  
RESULT 6  
US-09-912-484-2  
; Sequence 2, Application US/09912484  
; Patent No. 6358725  
; GENERAL INFORMATION:  
; APPLICANT: Christie, Gary  
; APPLICANT: Li, Xiaotong  
; APPLICANT: Powell, David J.  
; APPLICANT: Zhu, Yuan  
; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (mASp1)  
; FILE REFERENCE: GP-70663-D1  
; CURRENT APPLICATION NUMBER: US/09/912,484  
; CURRENT FILING DATE: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/166,974  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: 09/717,432  
; PRIOR FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: MUS MUSCULUS  
US-09-912-484-2

Query Match 47.4%; Score 1146.5; DB 4; Length 514;  
Best Local Similarity 50.7%; Pred. No. 1.1e-114;  
Matches 216; Conservative 75; Mismatches 130; Indels 5; Gaps 3;  
QY 8 BPGR-RGSFVEMVDNLKSGQGYVMTGSPQTLNILDVTSIPHPNVTVRANIAAITESDKFFI 126  
Db 65 EPRVATANFLAMVDNLQDSCRGYLEMLICTPPKQVQLVDTGSSNFVAVAGAPHSYIDT 124  
QY 67 TYQRLSTYRDLRGVTVPTQCKWEGELGTDLVSIPIHGPNTVVRANIAAITESDKFFI 126  
Db 125 YFDESSTYHSGFDVTVKVTQGSWTGFGEDLVTPKGFNSFLVNIATIFESNEFFL 184  
QY 127 NGSNNEGILGLAYAEIARPDSDLEPFDLSVKQTHVPLNLSLQCGAGFPNLSQEVLASV 186  
Db 185 PGIKWNGILGLAYALAKPSSLETFDLSVAQAQKIPDIFSMQMGAGLPVAGS---GTN 241  
QY 187 GSGMIIGDHLSTYGLSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSG 246  
Db 242 GGSVLGGTEPLSLYKDIWYTIKEWYQYIEILKLEGGONLNDREYNADRAIVDSG 301  
QY 247 TTNLRPKKVFEEAAKSIKAASSTKFFPDGFWLQGLVQWQAGTTPNIFPVISILMGE 306  
Db 302 TLLRLPQKVFVAVAVARTSLIPEFSDGFWTGAQACWTNSETPWAFPKISILYRDE 361  
QY 307 VTNQSFRTILPQOYLRPVEDVATSDCCYKFAISQSSTGTVMGAVMEGFVYVDFDRAOR 366  
Db 362 NASRSFRITILPOLYTOPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGFVYVDFDRAOR 420  
QY 367 RIGFAVSACHVHDEFTAAVEGFVTLDMEDCGYNIPODESTLMTIAYVMAAICALFML 426  
Db 421 RVGFVSPCAEIGTGVSEISGPFSTEDIASNCVPAQALNEPILMTIVSVYALMSVCGAAILL 480  
QY 427 PLCLMV 432  
Db 481 VLILL 486

RESULT 7  
US-08-007A-13  
; Sequence 13, Application US/08208007A  
; Patent No. 5501969  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/208,007A  
; FILING DATE: March 8, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: No. 5501969e  
; FILING DATE: No. 5501969e  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-208-007A-13

Query Match 12.4%; Score 299.5; DB 1; Length 396;  
Best Local Similarity 25.9%; Pred. No. 1.3e-23;  
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;  
QY 3 DEEPFEPGRGSGFVEMVDNLKSGQGYVMTGSPQTLNILDVTSIPHPNVTVRANIAA 58  
Db 63 DQSAKEP-----LNYLD-----MEYFGTISIGSPQNFVIFDTGSSNLWYPSVYCT 110  
QY 59 APHPFLHRYQRLSTYRDLRGVTVPTQCKWEGELGTDLVSIPIHGPNTVVRANIAA 118  
Db 111 SPACKTHSRFQSSQSYVQGSQSFQYGTGSLGIADQSV-EGLTIVGQGFGEV 169  
QY 119 TESDKFFINGNIGILGLAYAEIARPDSDLEPFDLSVKQTHVPLNLSLQCGAGFPN 178  
Db 170 TEPQGTFFVD-AEFDGILGLGYPSLA--VGGVTPVDFNNMAQ-----NLVDLPMSVYMSN 222  
QY 179 QSEVLASVGGSMIIGDHLSTYGLSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNY 238  
Db 223 PE--GGAGSELIFGGYDHSFSGSLNWPVTKQAYWQIALDNIQVGG--TVMFCSE--G 275  
QY 239 DKSIVDSGTTNLRPKKVFEEAAKSIKAASSTKFFPDGFWLQGLVQWQAGTTPNIPV 298  
Db 276 COAIVDTGSLITGSDRIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322  
QY 299 ISLYLMGEVTNQSFRTILPQOYLRPVEDVATSDCCYKFAISQSSTG-----SSG 346  
Db 323 VTFTING-----VPYTLSPYAY--TLILDVFDGMQFC-----SSGFOGLDIHPPAGP 366



QY 347 -TVMGAVIMEGFYVVFDRARRKRGFA 371  
Db 367 LWILGDFVFIQFYSVFDGRNVRVGLA 392

RESULT 8  
US-09-032-523-9  
; Sequence 9, Application US/09032523  
; Patent No. 6232454  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Gugler, Karl  
; APPLICANT: Baugh, Mariah  
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,523  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0479 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 396 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 181994

US-09-032-523-9  
Query Match 12.4%; Score 299.5; DB 4; Length 396;  
Best Local Similarity 25.9%; Pred. No. 1.3e-23;  
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;  
QY 3 DEEPEEPGRGSGFYVMDNLGRKSGGCGYVEMTVGSPPTLNILVDFTGSSNFAVGA-----58  
Db 63 DQSAKEP-----LNYLD-----MEYFGTISIGSPQNFVTFDGGSNLWVPSVYCT 110  
QY 59 APHPFLHRYQRLSSYRDLRGVYVYTGKWEGLGTDLYSIPHPGNVTVRANIAAI 118  
Db 111 SPACKTHSRFPQSSTYSQPGSFSIQYGTGSLSIGIGADQVSV-EGLTVVQQGGSV 169  
QY 119 TESDKFFINSNNEGILGLAYAEIARPDSDLEPFEDSLYKQTHVFNLFSLQLCAGFPLN 178  
Db 170 TEPQQTVD-AEFDGILGLGYPSLA--VCGVTPVFONMAQ-----NLVDLPMEFSVTWSSN 222

QY 179 QSEVLASVGGSMIIIGIDHSYTGSLWYTPIRREWYVEVIVRVEINGQDLKMDCKEYNY 238  
Db 223 PE---GGAGSELIFGSDHSFSGSLNWVPYTKQAYWQIALDNIQVGG--TVMECSE--G 275  
QY 239 DKSIVDSGTTNLRPKKVFEEAAVKAASSTKPKPDGFWLGEQLVCMQAGTTPWNIPV 298  
Db 276 CQAIIVDTGTSITGFSKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMPD 322  
QY 299 ISLYLMGEVTNQSFRTITLPOQLRPVEDVATSDQDCYKFAISQSSTG-----346  
Db 323 VTFTING-----VPYTLSPYAY--TLLDVFDGMOFC-----SSGFQGLDIHPGAP 366  
QY 347 -TVMGAVIMEGFYVVFDRARRKRGFA 371  
Db 367 LWILGDFVFIQFYSVFDGRNVRVGLA 392

RESULT 9  
US-08-208-007A-12  
; Sequence 12, Application US/08208007A  
; Patent No. 5501969  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CECHELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/208,007A  
FILING DATE: March 8, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: No. 5501969e  
FILING DATE: No. 5501969e  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-208-007A-12

Query Match 12.2%; Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. No. 3.8e-23;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;  
QY 30 YVEMTVGSPPTLNILVDFTGSSNFAVGAAPHPFL-----HRYQRLSSYRDLRGV 83  
Db 79 YXGEIGTGPQCTVTFDGTSSNLWVPSIHKCLDIACWIHHKYNDSKSTYVKNGTSF 138  
QY 84 VVPYTGKWEGLGTDLYSIP-----HGPNTVTVRANIAAITESDKFFINSNMEGI 134  
Db 139 DIHGGSLGSLYSDTIVSPQCQSSASSALGGVKVQVFGEATKQPGITFIAAKFDGI 198





QY 30 YVEMTVGSPQTLNVLDTGSSNFAYGAPHPFL-----HRYTORQLSSTYRDLRKG 82  
Db 120 YFGKIYLGTPPQETVLFDTGSSDFWV---PSIYCKSNACKNHQREDPRKSTFQNLGKP 176  
QY 83 YVYPTOGKWEGLGDLVPSHPGNVTVRANTAAITESDKFFINGSNWEGILGLAYAEI 142  
Db 177 LSIHYGTGSMQICGLYDTVTSN--IVDIQOTVGLSTQEPGDVFTTAEFGILGMAYPSL 234  
QY 143 ARPDLSLEPFDLSLVKQTHV-PNLFSLQLCGAGPLNQSEVLASVSGSMITGGIDHSLYT 201  
Db 235 A--SEYSIPVDNMNRHLVAQDLFSYMDRNG---QESMLT-----LGAIDFSYTT 281  
QY 202 GSLWTPTRRREWEYEVIIIVVEINGQDLKMD--CKEYNDKSIIVDSCTTNLRPKKVFEA 259  
Db 282 GSLHWVPVTVQYQWFTVDSVTISGVVACRGGC-----QAILDTGTSLKLVGSSDILN 335  
QY 260 AVKSTKAASSTKEKFDGFWLGE-QLVCWQAGTTPWNPVVISLYLMGEVNTNQSPRITLP 318  
Db 336 IQQAIGATQNO-----YGEFIDICDNLSPVTVVF-----EINGKMYPLT--P 376  
QY 319 QOYLPRVEDVATSDDCY---KFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFA 371  
Db 377 SAY-----TSQDQGFCTSGFQSENHSQKWILGDLGVFIREYYSVDFRANLNLVGLA 424

RESULT 15  
US-08-633-2  
; Sequence 2, Application US/08088633  
; Patent No. 5324660  
; GENERAL INFORMATION:  
; APPLICANT: Gleeson, Martin A  
; APPLICANT: Howard, Bradley D  
; TITLE OF INVENTION: Genes which influence pichia proteolytic  
; TITLE OF INVENTION: Activity, and Uses Therefor  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fitch, Even, Tabin & Flannery  
; STREET: 135 South LaSalle Street, Suite 900  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; APPLICATION NUMBER: US/08/088,633  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/678,916  
; FILING DATE: 01-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: 50848  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)552-1311  
; TELEFAX: (619)552-0095  
; TELETYPE: 20 6566 PATLAW CGO  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-633-2

Query Match 11.2%; Score 270; DB 1; Length 410;  
Best Local Similarity 25.7%; Pred. No. 2.1e-20;

Db 96 YFTEITLGSPPQSFVKVILDTGSSNLMWVPAECGSLACFLHTKYDHEASSTYKANGSEFAI 155  
QY 86 PYTGKWEGLGDLVPSHPGNVTVRANTAAITESDKFFINGSNWEGILGLAYAEIARP 145  
Db 156 QYSGSLEGYVSRDLTI--GDLVIPDQDFAEATSEPGLAFAFGKPDGILGLAYDSIS-V 212  
QY 146 DDLSEPFDSLVKQTHVNPVNFSLQLCGAGPLNQSEVLASVSGSMITGGIDHSLYTGSLW 205  
Db 213 NRIVPVYNAIKNLLDDPFA-----FYLGDSD--KSDGGEAFSGGIDEKYTGTEIT 264  
QY 206 YTPRREWEYEVIIIVVEINGQDLKMDCKEYNDKSIIVDSCTTNLRPKKVFEAAVKSK 265  
Db 265 WLPVRRKAYWE-----VKFEGIGLGEYATLEGHGAIDTGTSLIALPSGLAEILNAEIG 319  
QY 266 AASSTKEKFDGFWLGEQLVCWQAGTTPWNPVVISLYLMGEVNTNQSPRIT----- 315  
Db 320 AKKG-----WSGOYSVDCESRDS---LPDLTLNFG---YNTITAYDYTLEVSG 363  
QY 316 -----ILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGF 370  
Db 364 SCISAFTHMDPPEPVGPLA-----IIGDAFLRKYYISYIDIGHDAVGL 405  
QY 371 AVSA 374  
Db 406 AKAA 409

RESULT 14  
US-08-846-021A-8  
; Sequence 8, Application US/08846021A  
; Patent No. 5948682  
; GENERAL INFORMATION:  
; APPLICANT: Moloney, Maurice M.  
; TITLE OF INVENTION: Preparation of Heterologous Proteins on  
; TITLE OF INVENTION: Oil Bodies  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERESKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: April 25, 1997  
; APPLICATION NUMBER: US/08/846,021A  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 9369-039  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 361-1398  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 427 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-846-021A-8

Query Match 11.3%; Score 273.5; DB 2; Length 427;  
Best Local Similarity 26.4%; Pred. No. 9.4e-21;  
Matches 94; Conservative 68; Mismatches 129; Indels 65; Gaps 15;

```
Matches 91; Conservative 51; Mismatches 162; Indels 50; Gaps 10;
Qy 30 YVEMTVGSPQTLNILDVTGSSNFAVCA---APHPFLHRYQRLSSYRDLRKGVYV 85
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 96 YFEVSLCTPPQSFVKVILDTGSSNLWVPSKDCGSLACFLHAKYDHDSSYKKNSSFEI 155
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 86 PYTQKWEGLGTDLVSPHGPNTVVRANIAAITESDKFFINGSNWEGILGLAYAEIARP 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 RYGGSGMEGYVSQDVLIQI--GDLTPKVDFAEATSEPCLAFAFGKFDGILGLAY----- 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 146 DDSLEPPFDLSLVKQTHVPNLF---SLQLCGA---GFPLNQSEVLASVGGSMIIGDHS 199
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 -----DSISVKNIPFPYIKALELDLDEPKFAFYLGDTDKDESDGLATFGGVDKSK 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 200 YTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIIVDSGTTNLRPKVFEA 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 YEGKITWLPVRRKAYWE-----VSFDGVLGSEVAELQKTGAADTGTSLIALPSGLAEI 314
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 260 AVKSIIKAASSTEKFPDGFNLGEQLVCHQAGTTPWNIPFPVISLYLMGEVTNQSFRIILPQ 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 LNAEIGATKG-----WSGOYAVDCDTRDS----LPDLTLTFAG-----YNFTITPY 356
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 320 QYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGCVVVEDRARKRIGFAVS 373
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 DYTLEVSGSCISAFTPMDFP-EPIGPLAIIIGDSFLRKYYSVYDLGKDAVGLAKS 409
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: August 7, 2002, 09:14:48  
Job time: 73 sec

